Structure of Small Virus-like Particles Assembled from the L1 Protein of Human Papillomavirus 16

Xiaojiang S. Chen,*‖ Robert L. Garcea,‡
Ilya Goldberg,*† Gregory Casini,‡
and Stephen C. Harrison*†§
*Department of Molecular and Cellular Biology
†Howard Hughes Medical Institute
Harvard University
Cambridge, Massachusetts 02138
‡Section of Pediatric Hematology/Oncology
University of Colorado School of Medicine
Denver, Colorado 80262

Summary

The papillomavirus major late protein, L1, forms the
pentameric assembly unit of the viral shell. Recombi-
nant HPV16 L1 pentamers assemble in vitro into capsid-
like structures, and truncation of ten N-terminal residues
leads to a homogeneous preparation of 12-pentamer,
icosahedral particles. X-ray crystallographic analysis
of these particles at 3.5 Å resolution shows that L1
closely resembles VP1 from polyomaviruses. Surface
loops contain the sites of sequence variation among
HPV types and the locations of dominant neutralizing
epitopes. The ease with which small virus-like parti-
cles may be obtained from L1 expressed in E. coli
makes them attractive candidate components of a
papillomavirus vaccine. Their crystal structure also
provides a starting point for future vaccine design.

Introduction

Human papillomaviruses (HPVs) are pathogens of epi-
thelial surfaces associated with both benign warts and
malignant tumors (Howley, 1996; Shah and Howley,
1996). Over 90 HPV subtypes have been identified based
on DNA sequence relationships. Strong epidemiologic
and biochemical evidence supports association of infec-
tion by certain high-risk HPV subtypes (e.g., HPV16, 18,
and 31) with subsequent development of human cervical
cancer (Pisani et al., 1993; Bosch et al., 1995).

Papilloma virions contain two virally encoded pro-
teins, L1 and L2, synthesized late in the infectious cycle.
These two proteins encapsidate a histone-associated,
closed circular, double-stranded DNA minichromo-
some. The outer shell of the virion contains 72 pentamers
of L1, centered on the vertices of a T = 7 icosahedral
lattice (Baker et al., 1991; Trus et al., 1997). L2, a largely
internal protein, is present at about 1/30 the abundance
of L1 (Kirnbauer et al., 1993). The overall surface organi-
zation of the papillomavirus shell resembles that of its
close relatives, murine polyomavirus and SV40, which
also contain 72 pentamers of a principal capsid protein,

§ To whom correspondence should be addressed (e-mail: schadmin@
crystal.harvard.edu).
‖ Present address: Department of Biochemistry and Molecular Ge-
netics, University of Colorado School of Medicine, Denver, Colorado
80262.

VP1 (Rayment et al., 1982). Papillomavirus particles are
somewhat larger than polyoma virions (600 Å diameter,
rather than 500 Å), and they contain a correspondingly
larger genome (8 kb rather than 5 kb) (Howley, 1996).
Despite the close structural relationship, however, the
genomic organization is distinct, and there is no discern-
ible sequence similarity between L1 (about 500 residues)
and VP1 (about 370 residues) (Belnap et al., 1996).

Recombinant expression of L1 using vaccinia, bacu-
lovirus, or yeast systems results in the formation of virus-
like particles (VLPs), which are assemblies of roughly
72 L1 pentamers in a shell similar to that of the virion
(Zhou et al., 1991a; Kirnbauer et al., 1992; Hagensee et
al., 1993; Rose et al., 1993; Sasagawa et al., 1995). VLPs
can induce neutralizing antibodies in inoculated ani-
mals, and immunization with VLPs can protect experi-
mental animals from subsequent challenge with infec-
tious virus (Breitburd et al., 1995; Suzich et al., 1995;
Christensen et al., 1996). Thus, VLPs appear to be excel-
lent candidates for papillomavirus vaccines.

Structural analysis of papillomavirus particles has
been limited by an inability to grow large amounts of
virus in culture and by a lack of success in obtaining
suitable crystals of virions purified from warts. We have
expressed the HPV16 L1 protein in E. coli and deter-
mined its crystal structure at 3.5 Å resolution. The crys-
tallization conditions favor formation of a 12-pentamer,
T = 1 icosahedral assembly of L1, which we describe
as a "small VLP." Assembly experiments in vitro show
that a short, N-terminal segment of the L1 polypeptide
chain acts as a switch between 72-pentamer, virion-like
assemblies and small VLPs. The L1 subunits have similar
folded structures as their polyoma VP1 counterparts,
but the modes of interpentamer association in polyoma
virions and papilloma small VLPs are quite distinct. The
L1 structure can be used to direct engineering of chime-
ric VLP vaccines (Greenstone et al., 1998), locate domi-
nant neutralizing epitopes, identify mechanisms of neu-
tralization by antibodies, and rationalize differences
among viral serotypes.

Results and Discussion

Expression, Crystallization, and Assembly

We expressed HPV16 L1 and various truncated forms
in E. coli, as glutathione S-transferase (GST) fusion pro-
teins. Contrary to our expectations based on polyoma
and SV40 VP1 (Liddington et al., 1991), deletions of up
to 30 C-terminal residues had little or no effect on the
stability or solubility of the expressed protein, or on
the capacity of the appropriate N-terminally truncated
protein to crystallize; C-terminal deletions longer than
30 residues rendered the L1 extremely sensitive to pro-
tease. N-terminal truncation was tolerated up to at least
residue 13, but only products of constructs lacking 10
N-terminal residues (L1 ΔN-10) gave crystals suitable
for structure determination. L1's with N-terminal dele-
tions of 15 or more residues were unstable in the fusion

Table 1. HPV16 L1 Truncations and Their Assembly Properties

Deletion	Trypsin Sensitivity	Apparent Diameter of Assembled Particle (Å)[a]
ΔN=0	No	600
ΔN=8	No	600
ΔN=9	No	600
ΔN=10	No	300
ΔN=15	Yes[b]	NA
ΔN=20	Yes	NA
ΔC=16	No	600
ΔC=30	No	600
ΔC=46	Yes	NA
ΔC=86	Yes	NA

ΔN/C-x: Δ designates deletion, N or C designates N or C terminus, and x designates number of deleted residues. All the N-terminal deletions were confirmed by seven cycles of N-terminal sequencing using Edman degradation. All the C-terminal deletions were confirmed by DNA sequencing. Trypsin sensitivity was used a measure of correct folding. The protease treatment was carried out while the GST-fused protein was still bound to glutathione beads (see Experimental Procedures). Sensitivity was determined by SDS-PAGE analysis of the beads after digestion.
[a] Many of the 600 Å particles are imperfect, and there are particles of other sizes seen on the grid (Figure 1, left). The major species, however, is a shell with the approximate size of an HPV virion.
[b] Most of the protein with this deletion is sensitive to digestion by trypsin, but about 5% resists degradation.
NA, not applicable.

constructs we used (Table 1), and a form with 13 residues removed could not be cleaved from the GST.

We obtained several crystal forms of L1 ΔN-10, all with unit cells much larger than expected for isolated pentamers. The various buffers that yielded crystals had in all cases pH less than 6.4, and subsequent studies showed that low pH favors assembly of L1 into VLPs (Figure 1). The size and homogeneity of the observed particles depend on N-terminal truncation. L1 pentamers with up to nine N-terminal residues removed yield VLPs that resemble 72-pentamer virions (Figure 1, left). L1 pentamers with 10 residues removed yield 12-pentamer, small VLPs (Figure 1, center), and addition of

even a single glycine to the N terminus of the form lacking 10 residues switches assembly back to larger shells (Table 1). We were unable to test the effect of longer deletions on assembly, for reasons of expression yields outlined above and in Table 1. The first ordered residue in our structure of a small VLP is Lys-20, and we believe that had we been able to prepare suitable pentamers with up to about 19 N-terminal residues removed, we would probably have obtained T = 1 assemblies at low pH.

The presence of the N-terminal segment of L1 thus determines the diameter of the assembled particles. The structure described below shows that the N terminus of L1 indeed lies at interpentamer contacts. The assembly properties of L1 also determine successful crystallization. All crystals that we obtained were of the 12-pentamer particle. The 72-pentamer shells, formed by in vitro assembly of full-length L1, are too irregular for successful crystallization (see Figure 1).

Structure Determination

The structure was determined using a bloblike model to generate very low-resolution, "molecular replacement" phases, which were extended from 31 Å to about 17 Å by 20-fold noncrystallographic symmetry averaging. These phases were then used to locate mercury atoms in a derivative, and mercury single isomorphous replacement and noncrystallographic symmetry refinement were used together to complete the phasing at 3.5 Å. Additional details are given in the Experimental Procedures.

Description of the Structure

The T = 1 icosahedral assembly with 12 pentamers (Figure 2) has an outer diameter of 318 Å. Within each pentamer, the principal domains of the five L1 subunits associate intimately with each other, making a tightly linked ring. The less extensive, interpentamer contacts are formed by small, laterally projecting, elbow-like domains, which cluster around three-fold symmetry axes

Full length L1, T=7   ΔN-10, T=1 (in vitro assembly)   ΔN-10, T=1 (from crystal)
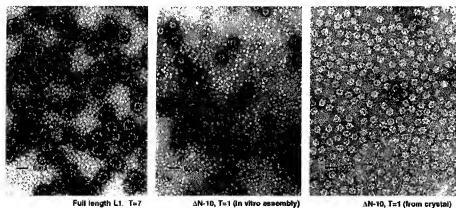
Figure 1. Assembly of Virus-like Particles from HPV16 L1 Pentamers, Monitored by Negative Stain Electron Microscopy
(Left) Particles that resemble 72-pentamer virions assemble at pH 5.2 from full-length L1. Note free pentamers in the background. (Center) Small, 12-pentamer VLPs assemble at pH 5.2 from L1 ΔN-10. Again, note free pentamers. (Right) Small VLPs derived from crystals, dissolved in low pH buffer. The staining is somewhat different than in the center panel because of the presence of some PEG, but close inspection shows the particles to be of indistinguishable morphology. Bars indicate 600 Å in all three panels.
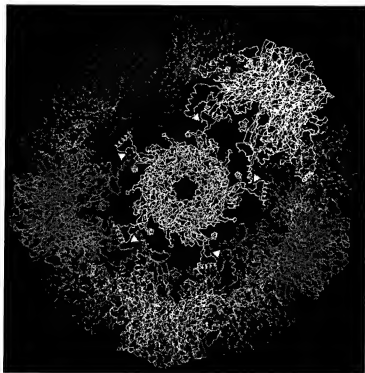
Figure 2. The T = 1 Particle of HPV16 L1

The view is near a five-fold axis. Each of the 12 identical L1 pentamers in this assembly is shown in a different color. The three-fold axes are shown by triangles. There are two-fold axes between each pair of three-fold axes.

(Figure 2). The complete particle thus has a markedly grooved appearance, with pentagonal "towers," decorated with stellate overhangs, rising above a more continuous "floor." The shape of these morphological units clearly corresponds to the pentamers seen by electron cryomicroscopy (Belnap et al., 1996; Trus et al., 1997).

The L1 Monomer

A classical "jelly roll" β sandwich, similar to the one found in polyoma and SV40 (Liddington et al., 1991), contains residues 20–382 of the 504-residue L1 polypeptide chain (Figures 3 and 4). The orientation of the long axis of the domain is approximately radial. Elaborate interstrand loops at one end impart a highly sculpted appearance to the outward-facing surface of the pentamer. The C-terminal lateral projections (residues 383–475, shown in yellow in Figure 3) are significantly α helical. Three helices (h2–4) form the surface of contact with other monomers, while a short strand (βJ) and a final helix (h5) anchor the projection back to the jelly roll. The strand adds to the C edge of the CHEF sheet, and the helix tucks into the axial cavity of the pentamer, where it has a hydrophobic interface with the base of the BIDG sheet. The last 31 residues are disordered and project toward the interior of the particle. This segment, the most variable among L1 proteins of different papillomaviruses, is rich in basic residues as well as serine and threonine; it probably interacts with DNA in the virion, and it includes the L1 nuclear localization signal (Zhou et al., 1991b). In most of our preparations, some of the disordered, C-terminal tails have been cleaved, suggesting that they are very susceptible to proteolysis before assembly. Recombinant L1, fused at its C terminus to another small protein and expressed in insect cells, can assemble into VLPs (Muller et al., 1997); it is likely that the fused element packs loosely into the interior of the VLP shell.

The Pentamer

The tightest contacts within a pentamer are at higher radii; the monomers splay apart toward the particle interior. The subunit orientation creates, along the pentamer axis, an inward-facing, conical hollow, which opens to the exterior of the particle through a narrow throat, about 14 Å in diameter. The polypeptide chain backbones of adjacent subunits interact directly. The G strand, at the inner margin of the BIDG sheet, "visits" the clockwise neighboring monomer and augments its CHEF sheet (see Figures 3A and 3C), before returning to rejoin its own subunit. The loops are also elaborately intertwined. The HI loop of one monomer, extending outward, inserts between the FG and EF loops of the anticlockwise neighbor and reaches far enough to contact the FG loop of the next neighbor; thus, there are, around the "top" of the pentamer, five overlapping bridges of paired HI:FG loops across an intervening subunit. Part of the EF loop extends outward to the edge of the pentamer, creating the five points of the starlike cap.

Pentamer–Pentamer Contacts

The only contacts between pentamers come from the small, laterally projecting domains (Figures 3A, 3C, and 3D). The first two helices (h2, h3) in the projecting domain form a V-shaped groove, into which fits the amino-terminal half of the longer, third helix (h4) from a three-fold related subunit. The contact is strongly hydrophobic (Figure 5). The third helix (h4) connects to the rest of the domain through an extended polypeptide chain, as if to create an adjustably oriented interaction
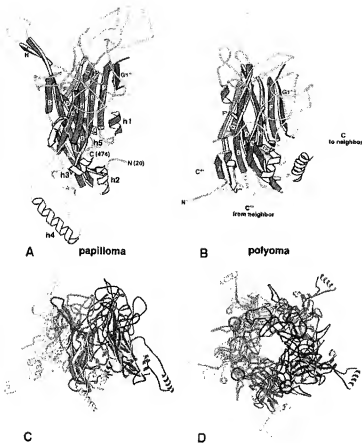
Figure 3. HPV16 L1 Monomer and Pentamer, with Polyomavirus VP1 Monomer for Comparison

(A) L1 monomer, viewed roughly normal to the five-fold axis of a pentamer. The structure includes residues 20–474; the first and last residues are labeled N(20) and C(474), respectively. The strands of the β jelly roll (and the short helix, h1, in the EF loop) are blue; connecting loops are pink; the G1 strand segment from the anticlockwise neighbor within a pentamer is red (this segment augments the CHEF sheet of the subunit, on the edge nearest the five-fold); the F strand of the clockwise neighbor is green (this is the edge of the sheet in the neighboring subunit with which the G1 segment forms hydrogen bonds); and the laterally projecting C-terminal domain is blue. Strands (capital letters) and helices (h1–h5) are labeled, and the corresponding sequences can be identified from Figure 3. A primed label (F') denotes a segment of polypeptide chain from the clockwise neighbor within the pentamer; a double-primed label (G1''), one from the anticlockwise neighbor.

(B) Polyomavirus VP1 monomer, viewed in essentially the same direction as the L1 monomer in (A). Blue, pink, red, and green have the same significance as in the representation of L1; a C-terminal arm, "invading" from another pentamer (triple-primed letters), is yellow; the proximal part of the arm of the monomer in the figure, which extends into the neighboring pentamer, is light yellow; distal parts of this arm are not shown.

(C) L1 pentamer, viewed normal to the five-fold axis. The two monomers in the rear have been deleted for clarity. The C-terminal projection of the central (blue) monomer is shown in yellow.

(D) L1 pentamer, viewed along the five-fold axis from outside the particle. As in (C), the C-terminal projection of the blue monomer is yellow.

element. In the virion, with 72 rather than 12 pentamers, the details of this contact may vary. The linkage of h4 is likely to be a locus of such flexibility.

The interpentamer contacts are modest, leaving significant gaps in the vicinity of the two-fold axes (Figure 2). In virions, and in 72-pentamer VLPs, the corresponding gaps may be filled by parts of the amino-terminal segments, which are partly missing and partly disordered in the small VLPs.

Comparison with VP1 of Polyomaviruses

The L1 subunit is clearly a close relative of VP1 of polyomaviruses, despite the absence of recognizable sequence similarity. Both form pentamers with radially oriented jelly roll domains; both link to each other through G strand interchange; both use C-terminal projections to form interpentamer contacts and to permit the six distinct packing environments for the subunit in the virion (Liddington et al., 1991). The outer surface loops of L1 are substantially longer than the corresponding loops of VP1—especially FG and HI, the two that make contact across an intervening monomer—but even the frameworks of the jelly roll domains are not readily superimposed. The two sheets formed by CHEF by BIDG of L1

lie more or less parallel to each other, while those of VP1 pack so that the β sandwich opens up toward the five-fold axis. Moreover, when the pentamer axes are aligned, the L1 jelly roll domain tilts away from the axis by 24° relative to VP1, so that the L1 pentamer is substantially wider at its base.

The most striking and unexpected difference between L1 and VP1 is the nature of the C-terminal, interpentamer linkages. The C-terminal arm of polyoma and SV40 VP1 is extended and flexible when the subunit is part of an unassembled pentamer (Stehle and Harrison, 1997; Chen et al., 1998). In the shell of the virion, each arm invades a neighboring pentamer, where it contributes an extra strand (βJ) to the β edge of the BIDG sheet of one subunit and forms additional, partly Ca²⁺-dependent contacts to another (Liddington et al., 1991; Stehle et al., 1994). Moreover, the βJ strand is clamped in place by a further strand (βA) from an N-terminal arm of the receiving subunit. Flexibility—the capacity to form a 72-pentamer structure—resides in the freedom of the arm to emerge from the jelly roll core in various directions and in the ability of the proximal residues of the arm to pack against other arms in somewhat variable ways. In the T = 1 particles seen in our crystals, the C-terminal

```
                                  27 β-B1      41 β-B2                        68 β-C
HPV16  HslWLPSeaTVVLPPv.PVGKVVsTbnTNIyTHAgtSRLLavGHPYIpikkpnnNKiiVPKVSglDPNKFgfPDLS
HPV18  HALWrPSdsTVVLPPss.VarVVnTDdYVtpTnIFYHAgSSRLLtvGHPfYirvpwgqqNEqdlPKVEAyQYKVPRvgLPDPNKFgLPDtS
HPV11  -snWrPSdsTVVtvPPpsPVGKVVsTDdYVkRTNIFYHAaSSRLLavGHPYYsfkkv..NKtvVPKVSgyQYRVPRVgsLPDPNKFALPDsS
CRPV  HAvWletgnKITLPP.qTVkLgsTDeYVkRTNvPYyAsSdRLLtTtGHPYYefr..dkgtnlVPKVSgmQYRVVFIkLPDPNKFAlgdDq
BPV1  HDLWqqgi.kiVgP.tPVSKVlcentYVqRtaIFTHAetaRLLlGHPYYpvsigakt...VPKVSsnQYRVVRlgLPDPmgFALPDrt
Cons  H--K--S--VYLPP--PV-KV--TD-YV-RT-IFYHA-S-RLLTVGHPYY---------VPKVS--QYRVFRI-LPDPNKF--PD--

        95                  P-D                                152     β-E
HPV16  iYNPdTQRLVWACvGVEVGRGQPLGVGiSGHPlLNKLDDtENAaaVAaNag..WDnRacimmDYKQTQLCiiGCkPpiGEHWGKGmpCbn
HPV18  iYNPeTQRLVWACsGVEILRGQPLGVGiSGHPfyNMLDDtEtyahaatNAvs..eDvRdnvavDYKQTQLCLiGCkPAiGEHWGKGtpCn
HPV11  iYdPtTQRLVWACtGiEVGRGQPLGVGvSGHPllNKyDDvENsggYqvNy..qDnRvnvgmDYKQTQLCmvGCkPPiGEHWGKGtqCsn
CRPV  iVdPdkeRiVWcLldiEVGRgGPLGVGVSGHPlLNKmrpgNhyRRkmatDpKEKVeHfvdakpgKKSGDe...i...
BPV1  vhRPskeRiLVgvplDVgVkRGQPLgiVtGHPtEkLIhaENvtvkvtt..qtDdmRkgtlDaKQgqliiLGCtPAeGEyWttargdvt
Cons  --P--RLVW---G-EV-RGQPLGV--SGHP--NK-DD-EN------N----D-RA------DYKQTQL--GC-P--GEHWGKG---C--

                                       234  H1     244       β-F
HPV16  vavnpGDCPPLELinTVIqpGDMVDTGFGAmDFsTLQAnRSEVPLDiCtSICKYPDYlKmvaePYGDSlFFYLRREQMFVRHlFNRAGAv
HPV18  rplaqGDCPPLELknTVIEDGDMVDTGYGAmDFsTLQAtRSEVPLDlCtSICKYPDYLqMsaDPYGDSMFFcLRREQlFARHFwkRAGTm
HPV11  tsvqnGDCPPLELLneTVIEDGDMVDTGFGAmDFsaLQssRSEVPLDLLmaDPYGDDiFYLAkEGMFARHFFNRAGTv
BPV1  diieniGECPPlkNkbLGDGMitLGFGAmnFatnaAkSDiPLDiqueiCiYPDyLkMaeDaaGndMPFtaNktvqvtvRtqGdae
CRPV  dgpqgtDCPPlELvnVrVqpGDlVDTGFGnm-LD--KGBVPLL----KCEVPDLqMamDPYGDr1FFYLRKEQMFARHFFNRAGKm
Cons  --DCPPLEL-N-VI-DGDM-D-GFGAm-F--LQ--KSEVPLD-----ICKYPDYL-M--D--GDSMFF--RREQMFARHFF-RAG--

           299 β-G1   310 β-G2          327 β-H1    342 β-H2
HPV16  GENVPdDLYIKGsgst....AnlASSnYfpTFSGSmVtSDaQLFNKPYWLQrAQGHNNGICWgNQLFVTVVDTTRSTNmsLCAmL.Stse
HPV18  GdtVPqelYIKGtgmp....AapgScvYpsaFSGSiVtSDsQIFNKPYWLhkAQGHNNGvCWhNQLFVTVVDTTRSTNLtLCAstdpvp
HPV11  GRpVPdDLvKGgmr......maWSSlYvhTPSGSLVSSeaQLFNKPYWLQKAQGHNNGvCWgNQLFVTVVDTTRSTNmTLCAsv.Sks
CRPV  kEmVKearNIkrgmagGnAaLAtohTclTPSGSLVSGSDpgvFNratWLqkAQGHNnGvNCWdGNQlFVTVVDnTGqTlLatvkkkeqik
BPV1  kRepttDY1iKnnkgaatlki...pDvhfgsFSGSLVStDnQlTNrPYWLfkAQGHNNGiaWgNqmPYGWGLKtvlpmtkCanvviPA..kkDp
Cons  G---W---YIK------------T-TPSGSLVSSD-Q-FNEPYWL-KAQG-HNG-CW-N-LFVTVVD-TR-TNLTL--------

          360       β-I      384  H2  395   H3          413      H4
HPV16  tTYknT.nFKeYlRHgEEYDLQFIFQLCkiTLTAdVMtYIHSMDstiLEPwNFGlgpPPqGTLEDTYRfVtSqAiaCQKhtPFApks.DP
HPV18  ggYdsT.kFKqYsRHvEEtDLQFIFQLCvIFLaAEVMSYIHSMNssiLEDwNFGVpPPPrtsinEDTYRyVQSqAitCQkttAPkakEDP
HPV11  sTYtns.dyKeYmRHvEEYDLQFIFQLCsiaLsAevMaYIHtMNpsvLEDwNFGlsPPPnGTLEDTYRyVQSqAitCQkgsPPkepkDe
BPV1  tefdkss.kFvrHVRHnEsYKLqgGCsaikLcKelsisTiMadiGAGtGAdnnnidLsSyslGhGnDtTDPYRfFwGtypktkpsvwssdgtp
BPV1  Com  --Y---F--Y-RH-EEY-LQFIFQLC---LSA---SYIHsM--S-LE-W-FG---PPP--TLEDtYRYV-S-A--C------P----

          446  β-I      462  H5
HPV16  lKkyrfWEVnLkEKFSaDLDQFPLGRKFLlQaCLkakpKflGRrkatPhtasSettaKRkrkl--
HPV18  YdkLRfWnVDLKEKFSlDLDQyPLGRKFLwQaSLrkptig.prKRaAPSattskpaKRvrvcark
HPV11  YKdmsFWEVnLKEEFSiKdglDQyPLGRKFLuQGyrqteartGlKRpAvRkp£tsRzKRtrk----
CRPV  YtknyKFWEVDLKEKidGgLDQyPLGRKFLuQGdl1q.....tlpKRpAPgpvSiVRsnKRtkrt----
BPV1  YegfKFWnLDLKEKidLDLDQyPLGRKFFaLeQGagcmiwkkrriagLtsBRpakkkkk--------
Cons  Y-----FW--V-LKEKFS-DLDQFPLGRKFL-Q-G---------------KR---------
```
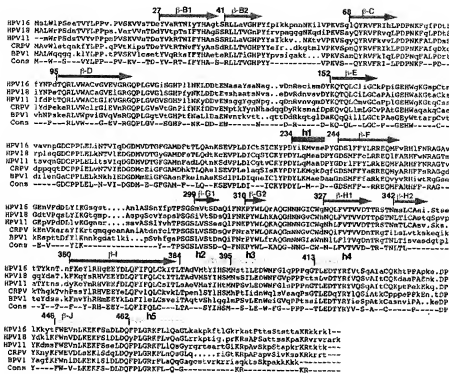
Figure 4. Amino Acid Sequences of Papillomavirus L1 Proteins
Alignment of L1 sequences from three human papillomavirus types (HPV16, 18, and 11), cottontail rabbit papillomavirus (CRPV), and bovine papillomavirus type 1 (BPV1) (PV website, 1997: http://hpv-web.lanl.gov). The secondary structural elements identified from the HPV16 L1 crystal structure are shown as arrows (strands) and rectangular bars (helices) in the line above the sequences. Blue, pink, and yellow colors have the same significance as in Figure 3A. The numbers refer to the sequence of HPV16.

part (residues 383–474) of the L1 polypeptide chain forms a projecting structure that presents acceptor and donor interaction surfaces, while remaining anchored to the jelly roll domain from which it emanates. Flexibility appears to reside in the extended strands that connect the donor surface (h4) to the rest of the subunit.

Control of L1 Assembly

We can in principle imagine that in the 72-pentamer virion, the C-terminal arms actually interchange, rather than return to their subunit of origin. In small VLPs, the arms of interacting pentamers approach each other near the N termini and the h4 helices (Figure 5B), and relatively minor local rearrangements could be sufficient to create an interchanged structure. The data presented in Table 1 indicate, however, that even in solution, the recombinant L1 pentamers we express in E. coli have anchored C-terminal arms, folded back as in the small VLP. C-terminal deletions that include any residues in helix h5, which anchors the arm, make the L1 extremely protease sensitive, indicating that an ordered h5 is important for stability (Table 1). Moreover, treatment of L1 pentamers with trypsin leads to a cleavage in the C-terminal projection, probably near one end of h4, but the short C-terminal fragment thus produced remains associated with the rest of the pentamer, presumably through tight interactions of h5 (data not shown). Full-length recombinant pentamers assemble

in vitro into virion-like shells (Figure 1, left). The conditions that lead to this assembly, which are the same as those that produce small VLPs from N-terminally truncated L1 (Figure 1, center), are unlikely to promote dissociation of the h5 anchor from its site on the subunit to which it belongs and insertion of the anchor at a related site on another pentamer. Thus, we conclude that assembly of virion-size particles can proceed without arm interchange. It is still possible that in vivo assembly in the presence of appropriate chaperones, which might protect exposed h5, does involve arm interchange. This mode of association would be a form of "domain swap," now documented to occur in many cases in addition to virus assembly (Schlunegger et al., 1997). This possibility can be tested by comparing properties of 72-pentamer particles assembled in vitro with those derived from expression in eukaryotic cells.

Why does deletion of ten N-terminal residues lead to assembly of a 12-pentamer rather than a 72-pentamer shell? The first ordered residue in our structure, Lys-20, lies near h2 in the C-terminal projection. The N terminus of the protein will thus be adjacent to the h3/h4 interpentamer contact, consistent with its observed effects. The amino-terminal, 20-residue segment is also a good candidate to form a structure that can fill interpentamer gaps in virions and VLPs (Figure 4). Unlike the largely basic, C-terminal disordered segment, the N-terminal region of L1 is relatively well conserved, both in length
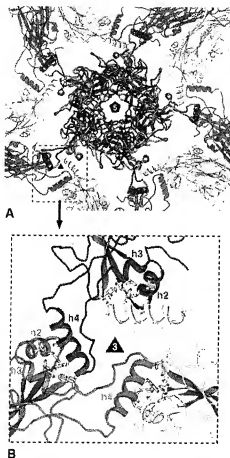
A

B

Figure 5. Five-Fold and Three-Fold Interactions in the HPV16 L1 T = 1 Particle

(A) View along a five-fold axis, showing how helix 4 from each subunit projects laterally and rests in a cradle formed by helices 2 and 3 of a subunit in an adjacent pentamer. All subunits in the central pentamer are red; two monomers in each adjacent pentamer are shown in yellow and green, respectively. The region surrounded by a dashed box is magnified in (B), where the contacts around a three-fold axis are displayed in detail. The strands of the β jelly roll in each monomer are blue; the C-terminal projections are red, yellow, and green, corresponding to the colors of the monomers in (A). The hydrophobic side chain contacts between helix 4 and the helix 2–helix 3 cradle are shown explicitly.

and in sequence, and the position of residue 20 is precisely consistent with a gap-filling function. In SV40 and polyoma virions, it is a β hairpin formed by a stretch of about 20 residues at the extreme C terminus of the polypeptide chain that has this role (Liddington et al., 1991).

Two cysteines, conserved among all papillomaviruses, have been reported to participate in VLP formation and virion stabilization, presumably through an interpentamer disulfide bond (Li et al., 1998; Sapp et al., 1998). These residues, 175 and 428 in HPV16, are not linked in our structure, and they are too distant from each other or from any cysteine in a neighboring pentamer to engage in disulfide bonds. In the larger, 72-pentamer shell, helix h4 (the loosely tethered contact

helix: Figure 5B) will have a somewhat different orientation than in the small VLP, and Cys-428 at its C-terminal end may be close enough to Cys-428 of a contacting subunit that the two can form an interpentamer disulfide bridge. A similar set of disulfides appears to occur in SV40, through a cysteine at the base of the CD loop.

L1 Variability

We have used a set of L1 sequences from 49 different HPV types (PV website, 1997: http://hpv-web.lanl.gov) to map the correlation of structure and sequence variation. Alignment shows that highly variable stretches are interspersed among segments of conserved residues. Display of conservation in three dimensions by color coding a model (Figure 6) reveals that all the hypervariable regions lie on the outward-facing surface of the pentamer. The lack of conservation may come simply from drift due to weak functional constraints, from negative selection by neutralizing antibodies or other components of an immune response, or from divergence to the use of different cell surface receptors. Because even the subunits' solvent-exposed lateral faces, which lie beneath the overhang of the star-shaped "cap," are far less variable than the loops on the top, we favor the notion that the changes have been fixed during evolution of the various HPV types by interaction with host functions. There is no evidence for tissue specificity of papillomavirus receptors (Roden et al., 1996a; Evander et al., 1997), and selection for escape from some component of the immune system appears to be the most likely interpretation of the observed pattern of variability.

Potential Receptor Sites

The outer surface of the L1 pentamer has five broad pockets, created by the BC, EF, and FG loops. While the rim of the L1 pocket is extremely variable, the floor is somewhat more conserved. These characteristics make the pockets likely candidates for receptor interaction, and indeed the receptor binding pockets on polyomavirus VP1 are in structurally homologous positions (Stehle et al., 1994; Stehle and Harrison, 1996).

Receptors for papillomaviruses have not been definitively identified. The presence of α6 integrin on the cell surface appears to be required for detectable HPB6b VLP binding, and it has been proposed that members of this class of integrins participate in HPB6b viral uptake (Evander et al., 1997). Functional receptors are present on a broad range of cell types, and bovine and human papillomaviruses can enter by the same pathway (Roden et al., 1996a). Despite their limited host range and tropism, papillomaviruses can bind to cells derived from a variety of tissues and from various species (Roden et al., 1994a; Muller et al., 1995; Volpers et al., 1995). Moreover, particles containing bovine papillomavirus type 1 (BPV1) genomic DNA encapsidated by L1/L2 from either HPV16 or BPV1 can infect the same mouse cell line (Roden et al., 1996a).

Neutralizing Epitopes

Antibody-mediated neutralization of papillomaviruses appears to have at least two distinct mechanisms. Some neutralizing antisera and monoclonal antibodies (mAbs) block cell binding, presumably by steric interference
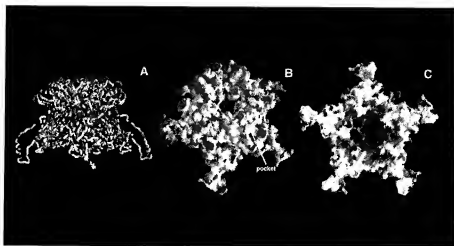
Figure 6. Distribution of L1 Sequence Variation in 49 HPV Types

Highly variable positions are red, fully conserved positions are blue, and positions of intermediate variation are white. (A) Wormlike representation, colored as described, of an L1 pentamer, viewed normal to the five-fold. (B) Surface representation of a tilted "side view" of an L1 pentamer. (C) Interior surface of the pentamer, showing conservation of residues (blue) in the conical cavity along the five-fold axis.

with the receptor site on the virus; others do not prevent attachment (and some even bind cell-attached VLPs) but probably inhibit uncoating (Roden et al., 1994b; Christensen et al., 1995; Booy et al., 1998). Structural analysis by electron cryomicroscopy of BPV1 decorated with two different neutralizing mAbs has yielded images consistent with these alternatives (Trus et al., 1997). One of the mAbs, which blocks attachment, binds to the outer surface of the pentamer; the other, which does not interfere with receptor interaction, binds across the groove between pentamers and presumably stabilizes the particle by cross-linking.

Neutralizing antibodies against papillomaviruses are highly type specific (Roden et al., 1996b). Epitopes have been mapped for neutralizing mAbs raised against HPV16 and HPV11 VLPs (Ludmerer et al., 1996, 1997; Roden et al., 1997; White et al., 1999), and Figure 7A shows positions on the pentamer surface where mutations in L1 influence neutralizing-antibody binding. The site for the V5 epitope of HPV16 lies near the center of the putative receptor pocket, and attachment inhibition is a likely neutralization mechanism. It is possible, however, that an IgG bound at this position could bridge between pentamers in a virion, where the radial divergence of pentamers is less marked than in the small, T = 1 particle. The V5 site is a dominant neutralizing epitope on HPV16; the corresponding mAb competes with about 75% of all patient antisera (White et al., 1999).

Within a principal papillomavirus subtype, variants arise with minor differences in L1 sequence; these isolates can differ in their affinity for particular neutralizing mAbs (Roden et al., 1997). The locations of sequence differences in 85 analyzed variants of HPV16 (Figure 7B) are all close to the principal neutralizing epitopes (Figure 7A), consistent with the notion that fixation of these sequence variants is driven by escape from neutralization by antibodies in the host. A distinct pattern of mutations has been observed in 57 variants of HPV5 (Figure

7C), suggesting that there may be a rather different set of dominant neutralization epitopes for different HPV types.

L2 Interactions

Like VP2/3 of the polyomaviruses, L2 probably interacts with an L1 pentamer in the conical hollow that faces inward along the five-fold axis (Griffith et al., 1992). The crystal structure of a complex of a VP1 pentamer and a fragment of VP2/3 shows that the common C-terminal segment of VP2/3 inserts in a looplike fashion into the axial cavity, so that both the N-terminal part of VP2/3 and the very C-terminal part extend toward the interior of the virion (Chen et al., 1998). Side chains that line the corresponding axial cavity in L1 are almost totally conserved, and even the narrow "throat" through which the hollow connects to the exterior contains invariant residues (Figures 6B and 6C). The amino-terminal half of the L2 polypeptide chain is likewise quite conserved. Both extreme amino- and carboxyl termini of L2 are rich in positively charged residues, and it is plausible that both interact with the viral genome. MAbs to certain L2 epitopes bind L1/L2 VLPs, however, and some of these mAbs are neutralizing (Campo et al., 1997; Kawana et al., 1999). The neutralizing epitope has in one case been mapped to the vicinity of L2 residue 108 (Kawana et al., 1999). This segment of L2 might project as a loop from the viral shell, either through fenestrations between pentamers or through the throat along the pentamer five-fold axis. Unlike polyomaviruses, where one internal protein associates with each VP1 pentamer, the ratio of L2 chains to L1 pentamers is distinctly substoichiometric. One suggestion is that only the 5-coordinated L1 pentamers contain L2 (Trus et al., 1997), but a more random distribution is also possible.

Vaccine Strategies

VLP-based vaccines are effective in preventing papillomavirus infection in animal models (Breitburd et al.,
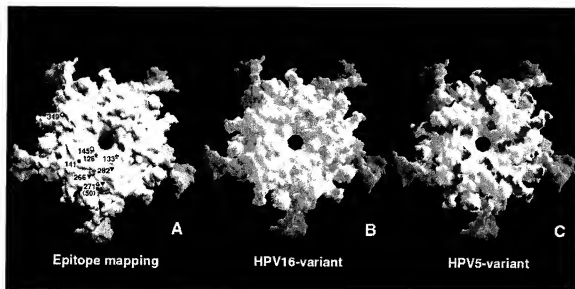
Figure 7. Aspects of the Antigenic Structure of L1

(A) Positions of neutralization epitopes in HPV16 (blue labels) and HPV11 (red labels), shown schematically on a surface representation of the pentamer, viewed along the five-fold axis from the "top." Only one set of five-fold related sites is shown. Open triangle, H16.V5; closed triangle, H16.E70; star, H11.B2, H11.F1, and H11.G5; closed square, H11.B2 and H11.F1; open square, H11.F1 and H11.H3. Residue 50, which was shown to affect HPV16 L1 binding to H16.E70 and H16.V5, is situated beneath the outer surface near residue 271 of the FG loop and is likely to influence the loop structure centered on this residue (Ludmerer et al., 1996, 1997; Roden et al., 1997; White et al., 1999).
(B) Positions of sequence variation among HPV16 isolates, shown as yellow patches projected onto the outer surface of the L1 pentamer.
(C) Positions of sequence variation among HPV5 isolates, shown as red patches on the L1 outer surface.

1995; Suzich et al., 1995; Christensen et al., 1996), and VLP immunogens prepared from yeast and insect cells are now in phase II clinical trials. The L1 structure we have described has several implications for vaccine design. First, the T = 1 particle itself—in effect, a small VLP—may be a suitable antigen for administration as a

Table 2. Data Collection, Phase Determination, and Refinement Statistics

Data Set	Space Group	a (Å)	Resolution	Completeness	R_{merge}
Native	$P2_13$	387.0	70–3.5 Å	87.5%	12.0%
			(3.62–3.5 Å)	(56.0%)	(41.0%)
Mercury	$P2_13$	388.2	70–5.5 Å	88.0%	12.1%
			(5.7–5.5 Å)	(75.0%)	(35.6%)

	Phasing from Initial Model Particle			CH_3Hg^+ SIR Averaging	Phase Extension
	Before Averaging	After Averaging	Phase Extension		
	70–31 Å	70–31 Å	70–17 Å	25–5.5 Å	25–3.5 Å
Correlation coefficient	37.3%	81%	77%	95.1%	91.2%
R_{avg}	60%	39%	35%	16.9%	25.7%

Refinement R Factors	No Refinement	Final		
		R_{work}	R_{free}	
15–3.5 Å	42%	28.3%	29.1%	
(3.6–3.5 Å)		(35.2%)	(36.5%)	

$R_{merge} = \Sigma|I - <I>|/\Sigma I$, where I is the observed intensity, and $<I>$ is the weighted average intensity of multiple observations of symmetry-related reflections. The sums are over all independent reflections for which there are multiple observations.
$R_{avg} = \text{sqrt } \Sigma[[w(F_o - F_c)^2]/[w(F_o)^2]]$, where F_c and F_o are calculated and observed structure factors, and w is a weight for each reflection over which the sum is taken.
R_{work} (R_{free}) $= \Sigma|F_o - F_c|/\Sigma|F_o|$, where F_c and F_o are the calculated and observed structure factors, and the sums are over all reflections.
Correlation coefficient $= \Sigma_h(|F_o| - <F_o>) (|F_c| - <F_c>)/[\Sigma_h (|F_o| - <F_o>)^2 \Sigma_h(|F_c| - <F_c>)^2]^{1/2}$, where F_o and F_c are observed and calculated structure factors, and the sums are over all reflections.
Numbers in parentheses refer to the outermost resolution bin.

vaccine. Presentation of epitopes on individual penta-mers is likely to be identical in the two contexts. The higher curvature of a small VLP implies that the pen-tamer tips are somewhat more widely separated than in a 72-pentamer shell, but the distances suggest that IgGs could still form bridges, if such interactions were important for B cell receptor affinity. Second, the analysis of epitopes and variants in Figure 7 provides a rational approach to cross-protective VLP (or T = 1) vaccines, by grafting segments of other HPV types—or even of other proteins—onto the same L1 matrix (Chackerian et al., 1999). Previous efforts to do so have involved effectively random trial and error. Third, the structure suggests how new interpentamer interactions could be introduced to stabilize the T = 1 particles, and through model-building, the 72-pentamer VLPs as well. Efforts to obtain the structure of a 72-pentamer particle, by fitting the results of high-resolution electron cryomicroscopy (Trus et al., 1997) and perhaps by crystallizing VLPs, are clearly relevant in this regard. Finally, by analogy with our previous work on polyomavirus VP1/VP2 complexes (Chen et al., 1998), we can now devise strategies for determining the location of L2 neutralizing epitopes, through studies of appropriate coassemblies.

Experimental Procedures

Protein Preparation and Crystallization

Recombinant HPV16 L1 was expressed in *E. coli* as a GST fusion protein, using the pGEX-2T expression vector, and purified from the supernatant of disrupted cells by glutathione-Sepharose chromatography. As described in greater detail elsewhere (X. S. C. et al., unpublished data), it was necessary to add ATP and 3.5 M urea to the cell lysate to release tightly bound GroEL. The GST moiety was removed by thrombin cleavage; subsequent Superdex chromatography yielded pure L1 pentamers. During crystallization trials, only the deletion mutant lacking the N-terminal ten amino acids yielded good crystals in the presence of 7.5% PEG in 0.2 M NaOAc buffer (pH 4.5). We obtained several crystal forms; the space group of the form used to determine the structure was P2₁3, a = 387 Å.

In Vitro Assembly

Purified pentamers produced by the corresponding deletion constructs (Table 1) were diluted from 10 mg/ml to 10 μg/ml with buffer containing 0.2 M Na acetate (pH 5.2), 1 M NaCl and allowed to stand at room temperature for 30 min. Samples were then spotted onto glow-discharged, carbon-coated grids and stained with 2% uranyl acetate for electron microscopy with a JEOL 100-CX microscope, operated at 80 kV.

Data Collection

Both native and mercury derivative (methyl mercury hydroxide) data sets were recorded at the CHESS F1 beamline with the ACSD Quantum-4 ccd detector. Native data extended to 3.5 Å; derivative data, to 5.5 Å. The data were indexed and integrated with DENZO (Otwinowski and Minor, 1997) and further processed with SCALEPACK and CCP4 programs (Collaborative Computational Project, 1994; version ccp4_3.3). Statistics are in Table 2. The mean I/σ(I) in the last resolution bin is about 2.

Phase Determination

Initial, very low resolution phases were obtained by molecular replacement with an approximate model, in which the polyomavirus VP1 pentamer was used as a five-fold symmetric "lump" of about the correct size. Self-rotation functions were used to determine the particle orientation (a single parameter in this space group), and a single-parameter position search along the crystallographic three-fold was used to find the particle center. Two internal parameters

of the particle (pentamer azimuth and radius) were refined against the observed data, using structure–factor correlation as a target, before the initial phase calculation. Phases at very low resolution (70–31 Å) were refined and extended to 17 Å by 20-fold noncrystallographic symmetry (NCS) averaging using RAVE (Kleywegt and Jones, 1994). This phase set was adequate for locating the 120 mercury sites in a crystallographic asymmetric unit. Further 20-fold NCS averaging and phase extension based on the mercury SIR phases yielded a 3.5 Å resolution map with clear side chain features, into which the L1 model was built using the program O (Jones et al., 1991). Details of the structure determination will be published elsewhere (Chen and Harrison, unpublished data).

Refinement

The initial model had an R factor of 42% (8–3.5 Å). Using 20-fold NCS restraints throughout the calculations, 300 steps of positional refinement in CNS (Brünger et al., 1998) gave $R_{work} = 34.5\%$ and $R_{free} = 35.8\%$. Torsion-angle molecular dynamics refinement lowered these figures to 33.4% and 34.9%, respectively. Further thermal-parameter (B) refinement together with positional refinement using data between 15 and 3.5 Å yielded $R_{work} = 28.3\%$ and $R_{free} = 29.1\%$ (see Table 2). All residues of the final model have phi/psi angles in the allowed regions of a Ramachandran plot (PROCHECK; Laskowski et al., 1993), with 72% in the most favored region.

Illustrations

The figures were made using O (Jones et al., 1991) (Figure 2), Molscript (Kraulis, 1991) (Figure 3), Ribbons (Carson, 1987) (Figure 5), and GRASP (Nicholls et al., 1991) (Figures 6 and 7).

References

Baker, T.S., Newcomb, W.W., Olson, N.H., Cowsert, L.M., Olson, C., and Brown, J.C. (1991). Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophys. J. *60*, 1445–1456.

Belnap, D.M., Olson, N.H., Cladel, N.M., Newcomb, W.W., Brown, J.C., Kreider, J.W., Christensen, N.D., and Baker, T.S. (1996). Conserved features in papillomavirus and polyomavirus capsids. J. Mol. Biol. *259*, 249–263.

Booy, F.P., Roden, R.B., Greenstone, H.L., Schiller, J.T., and Trus, B.L. (1998). Two antibodies that neutralize papillomavirus by different mechanisms show distinct binding patterns at 13 Å resolution. J. Mol. Biol. *281*, 95–106.

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., and Shah, K.V. (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J. Natl. Cancer Inst. *87*, 796–802.

Breitburd, F., Kirnbauer, R., Hubbert, N.L., Nonnenmacher, B., Trin-Dinh-Desmarquet, C., Orth, G., Schiller, J.T., and Lowy, D.R. (1995). Immunization with viruslike particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. J. Virol. *69*, 3959–3963.

Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new

software suite for macromolecular structure determination. Acta Crystallogr. D 54, 901–921.

Campo, M.S., O'Neil, B.W., Grindlay, G.J., Curtis, F., Knowles, G., and Chandrachud, L. (1997). A peptide encoding a B-cell epitope from the N-terminus of the capsid protein L2 of bovine papillomavirus-4 prevents disease. Virology 234, 261–266.

Carson, M. (1987). Ribbon models of macromolecules. J. Mol. Graph. 5, 103–106.

Chackerian, B., Lowy, D.R., and Schiller, J.T. (1999). Induction of autoantibodies to mouse CCR5 with recombinant papillomavirus particles. Proc. Natl. Acad. Sci. USA 96, 2373–2378.

Chen, X.S., Stehle, T., and Harrison, S.C. (1998). Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. EMBO J. 17, 3233–3240.

Christensen, N.D., Cladel, N.M., and Reed, C.A. (1995). Postattachment neutralization of papillomaviruses by monoclonal and polyclonal antibodies. Virology 207, 136–142.

Christensen, N.D., Reed, C.A., Cladel, N.M., Han, R., and Kreider, J.W. (1996). Immunization with viruslike particles induces long-term protection of rabbits against challenge with cottontail rabbit papillomavirus. J. Virol. 70, 960–965.

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for crystallography. Acta Crystallogr. D 50, 760–763.

Evander, M., Frazer, I.H., Payne, E., Qi, Y.M., Hengst, K., and McMillan, N.A. (1997). Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. J. Virol. 71, 2449–2456.

Greenstone, H.L., Nieland, J.D., deVisser, K.E., De Bruijn, M.L., Kimbauer, R., Roden, R.B., Lowy, D.R., Kast, W.M., and Schiller, J.T. (1998). Chimeric papillomavirus-like particles elicit antitumor immunity against the E7 oncoprotein in an HPV16 tumor model. Proc. Natl. Acad. Sci. USA 95, 1800–1805.

Griffith, J.P., Griffith, D.L., Rayment, I., Murakami, T., and Caspar, D.L.D. (1992). Inside polyomavirus at 25-Å resolution. Nature 355, 652–654.

Hagensee, M.E., Yaegashi, N., and Galloway, D.A. (1993). Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. J. Virol. 67, 315–322.

Howley, P. (1996). Papillomavirinae: the Viruses and Their Replication (New York: Raven Press).

Jones, T.A., Zhou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47, 110–119.

Kawana, K., Yoshikawa, H., Taketani, Y., Yoshike, K., and Kanda, T. (1999). Common neutralization epitopes in minor capsid protein L2 of human papillomavirus types 16 and 6. J. Virol. 73, 6188–6190.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D.R., and Schiller, J.T. (1992). Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc. Natl. Acad. Sci. USA 89, 12180–12184.

Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Durst, M., Gissmann, L., Lowy, D.R., and Schiller, J.T. (1993). Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 virus-like particles. J. Virol. 67, 6929–6936.

Kleywegt, G.J., and Jones, T.A. (1994). From first map to final model. In Proceedings of the CCP4 Study Weekend, S. Bailey, R. Hubbard, and D. Waller, eds. (Daresbury, UK: Daresbury Laboratory), pp. 59–66.

Kraulis, P.E. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

Li, M., Beard, P., Estes, P.A., Lyon, M.K., and Garcea, R.L. (1998). Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly. J. Virol. 72, 2160–2167.

Liddington, R.C., Yan, Y., Zhao, H.C., Sahli, R., Benjamin, T.L., and

Harrison, S.C. (1991). Structure of simian virus 40 at 3.8 Å resolution. Nature 354, 278–284.

Ludmerer, S.W., Benincasa, D., and Mark, G.E., III. (1996). Two amino acid residues confer type specificity to a neutralizing, conformationally dependent epitope on human papillomavirus type 11. J. Virol. 70, 4791–4794.

Ludmerer, S.W., Benincasa, D., Mark, G.E., III, and Christensen, N.D. (1997). A neutralizing epitope of human papillomavirus type 11 is principally described by a continuous set of residues which overlap a distinct linear, surface-exposed epitope. J. Virol. 71, 3834–3839.

Muller, M., Gissmann, L., Cristiano, R.J., Sun, X.Y., Frazer, I.H., Jenson, A.B., Alonso, A., Zentgraf, H., and Zhou, J. (1995). Papillomavirus capsid binding and uptake by cells from different tissues and species. J. Virol. 69, 948–954.

Muller, M., Zhou, J., Reed, T.D., Rittmuller, C., Burger, A., Gabelsberger, J., Braspenning, J., and Gissmann, L. (1997). Chimeric papillomavirus-like particles. Virology 234, 93–111.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11, 281–296.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.

Pisani, P., Parkin, D.M., and Ferlay, J. (1993). Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. Int. J. Cancer 55, 891–903.

Rayment, I., Baker, T.S., Caspar, D.L.D., and Murakami, W.T. (1982). Polyoma virus capsid structure at 22.5 Å resolution. Nature 295, 110–115.

Roden, R.B., Kirnbauer, R., Jenson, A.B., Lowy, D.R., and Schiller, J.T. (1994a). Interaction of papillomaviruses with the cell surface. J. Virol. 68, 7260–7266.

Roden, R.B., Weissinger, E.M., Henderson, D.W., Booy, F., Kirnbauer, R., Mushinski, J.F., Lowy, D.R., and Schiller, J.T. (1994b). Neutralization of bovine papillomavirus by antibodies to L1 and L2 capsid proteins. J. Virol. 68, 7570–7574.

Roden, R.B., Greenstone, H.L., Kirnbauer, R., Booy, F.P., Jessie, J., Lowy, D.R., and Schiller, J.T. (1996a). In vitro generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype. J. Virol. 70, 5875–5883.

Roden, R.B., Hubbert, N.L., Kirnbauer, R., Christensen, N.D, Lowy, D.R., and Schiller, J.T. (1996b). Assessment of the serological relatedness of genital human papillomaviruses by hemagglutination inhibition. J. Virol. 70, 3298–3301.

Roden, R.B., Armstrong, A., Haderer, P., Christensen, N.D., Hubbert, N.L., Lowy, D.R., Schiller, J.T., and Kirnbauer, R. (1997). Characterization of a human papillomavirus type 16 variant-dependent neutralizing epitope. J. Virol. 71, 6247–6252.

Rose, R.C., Bonnez, W., Reichman, R.C., and Garcea, R.L. (1993). Expression of human papillomavirus type 11 L1 protein in insect cells: in vivo and in vitro assembly of viruslike particles. J. Virol. 67, 1936–1944.

Sapp, M., Fligge, C., Petzak, I., Harris, J.R., and Streeck, R.E. (1998). Papillomavirus assembly requires trimerization of the major capsid protein by disulfides between two highly conserved cysteines. J. Virol. 72, 6186–6189.

Sasagawa, T., Pushko, P., Steers, G., Gschmeissner, S.E., Hajibagheri, M.A., Finch, J., Crawford, L., and Tommasino, M. (1995). Synthesis and assembly of virus-like particles of human papillomaviruses type 6 and type 16 in fission yeast Schizosaccharomyces pombe. Virology 206, 126–135.

Schlunegger, M.P., Bennett, M.J., and Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv. Prot. Chem. 50, 61–122.

Shah, K.V., and Howley, P. (1996). Papillomaviruses (New York: Raven Press).

Stehle, T., Yan, Y., Benjamin, T.L., and Harrison, S.C. (1994). Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. Nature 369, 160–163.

Stehle, T., and Harrison, S.C. (1996). Crystal structures of murine

polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. Structure *4*, 183–194.

Stehle, T., and Harrison, S.C. (1997). High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J. *16*, 5139–5148.

Suzich, J.A., Ghim, S.J., Palmer-Hill, F.J., White, W.I., Tamura, J.K., Bell, J.A., Newsome, J.A., Jenson, A.B., and Schlegel, R. (1995). Systemic immunization with papillomavirus L1 protein completely prevents the development of viral mucosal papillomas. Proc. Natl. Acad. Sci. USA *92*, 11553–11557.

Trus, B.L., Roden, R.B., Greenstone, H.L., Vrhel, M., Schiller, J.T., and Booy, F.P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution. Nat. Struct. Biol. *4*, 413–420.

Volpers, C., Unckell, F., Schirmacher, P., Streeck, R.E., and Sapp, M. (1995). Binding and internalization of human papillomavirus type 33 virus-like particles by eukaryotic cells. J. Virol. *69*, 3258–3264.

White, W.I., Wilson, S.D., Palmer-Hill, F.J., Woods, R.M., Ghim, S.J., Hewitt, L.A., Goldman, D.M., Burke, S.J., Jenson, A.B., Koenig, S., and Suzich, J.A. (1999). Characterization of a major neutralizing epitope on human papillomavirus type 16 L1. J. Virol. *73*, 4882–4889.

Zhou, J., Sun, X.Y., Stenzel, D.J., and Frazer, I.H. (1991a). Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. Virology *185*, 251–257.

Zhou, J., Doorbar, J., Sun, X.Y., Crawford, L.V., McLean, C.S., and Frazer, I.H. (1991b). Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. Virology *185*, 625–632.

Protein Data Bank ID Code

The ID code for the coordinates of the structure reported in this article is 1dzl.

JMB

Papillomavirus Capsid Protein Expression in *Escherichia coli*: Purification and Assembly of HPV11 and HPV16 L1

Xiaojiang S. Chen[1]*, Gregory Casini[2], Stephen C. Harrison[3] and Robert L. Garcea[2]

[1]Department of Biochemistry & Molecular Genetics, University of Colorado School of Medicine, 4200 E. 9th Ave, Denver CO 80262, USA

[2]Section of Pediatric Hematology/Oncology University of Colorado School of Medicine, 4200 E. 9th Ave Denver, CO 80262, USA

[3]Harvard University and Howard Hughes Medical Institute, Department of Molecular and Cellular Biology 7 Divinity Ave, Cambridge MA 02138, USA

*Corresponding author

The L1 major capsid proteins of human papillomavirus (HPV) types 11 and 16 were purified and analyzed for structural integrity and *in vitro* self-assembly. Proteins were expressed in *Escherichia coli* as glutathione-*S*-transferase-L1 (GST-L1) fusions and purified to near homogeneity as pentamers (equivalent to viral capsomers), after thrombin cleavage from the GST moiety and removal of tightly associated GroEL protein. Sequences at the amino and carboxy termini contributing to formation of L1 pentamers and to *in vitro* capsid assembly were identified by deletion analysis. For both HPV11 and HPV16 L1, up to at least ten residues could be deleted from the amino terminus (ΔN10) and 30 residues from the carboxy terminus (ΔC30) without affecting pentamer formation. The HPV16 pentamers assembled into relatively regular, 72-pentamer shells ("virus-like particles" or VLPs) at low pH, with the exception of HPV16 L1 ΔN10, which assembled into a 12-pentamer, $T = 1$ capsid (small VLP) under all conditions tested. The production of large quantities of assembly-competent L1, using the expression and purification protocol described here, has been useful for crystallographic analysis, and will be valuable for studies of virus-receptor interactions and potentially for vaccine design.

© 2001 Academic Press

Keywords: human papillomavirus; virus assembly; virus capsids; protein folding

Introduction

The Papovaviridae include two virus subfamilies: papilloma- and polyomaviruses. Polyomaviruses are DNA tumor viruses that transform cells in tissue culture and cause tumors in experimental animals. Papillomaviruses are important human pathogens, whose infection is directly related to the subsequent development of cervical cancer for high-risk subtypes such as human papillomavirus (HPV) 16 and 18 (Bosch et al., 1995; Pisani et al., 1993).

Both papilloma- and polyomavirus capsids have a similar overall architecture: 360 identical copies of their major capsid protein are arranged as 72

pentamers at the vertices of a $T = 7$ icosahedral lattice (Rayment et al., 1982; Baker et al., 1991). Surface feature details, as determined by cryoelectron microscopic image reconstructions of virions, differ noticeably, however (Belnap et al., 1996; Trus et al., 1997). Comparison of VP1 proteins among different polyomaviruses (e.g. SV40, murine polyoma) and L1 proteins among different papillomaviruses (e.g. bovine papilloma, HPV) shows that the primary sequences are clearly related within the two subfamilies, but that there is no detectable sequence similarity between L1s and VP1s.

The crystal structures of the VP1 major capsid proteins from murine polyomavirus and SV40 are known (Liddington et al., 1991; Stehle et al., 1994). In addition, the biochemistry of *in vitro* capsid assembly for recombinant VP1 capsomeres has been studied in detail (Garcea et al., 1987; Salunke et al., 1989). Polyoma VP1 contains a jellyroll β-barrel core, and outside the jellyroll core there are long and flexible arms at both amino and

carboxy termini. These arms are dispensable for pentamer formation, but parts of both are required for inter-pentamer interactions and thus essential for particle assembly. The positively charged N-terminal segment is a DNA-binding domain and likely interacts with the viral DNA. The domain organization of HPV L1 domain differs from that of polyoma VP1 in that the positively charged segment is at the extreme carboxy terminus (Li *et al.*, 1997; Zhou *et al.*, 1991), following a sequence of 70 residues important for inter-pentameric interactions (Chen *et al.*, 2000).

Papillomaviruses are difficult to grow in cell culture, but the L1 proteins have been produced by recombinant expression in *Escherichia coli* and in insect cells using baculovirus vectors (Kirnbauer *et al.*, 1992; Le Cann *et al.*, 1994; Li *et al.*, 1997; Rose *et al.*, 1993; Xi & Banks, 1991). We report here a purification strategy for obtaining large quantities of pure, well-folded HPV11 and HPV16 L1 pentamers after recombinant expression in *E. coli*. Using deletion mutagenesis and purification of recombinant mutant L1 proteins, sequences at the amino and carboxy termini of L1 important for L1 folding and pentamer formation were identified. The carboxy-terminal sequences were important for particle assembly, and the amino terminus appeared to regulate the size of the icosahedral shell.

Results

Protein expression in *E. coli*

The L1 proteins from both HPV11 and HPV16 were analyzed in parallel in order to generalize the findings, as well as to utilize minor differences in the primary sequences to define domain boundaries. Both proteins behaved similarly, and only the differences are specifically discussed. The L1 coding sequences were cloned into pUC based vectors under the Ptac promoter for expression in *E. coli*. Using these constructs, the L1 protein was not detected by SDS-PAGE in whole cell lysates, despite efforts to improve expression by varying the DNA sequence near the L1 start codon, mutating several rare arginine codons, or changing bacterial strains (data not shown). The low level of L1 accumulation was subsequently determined to be due, in part, to protein degradation rather than low expression levels (data not shown). As an alternative approach, we fused glutathione-*S*-transferase (GST) to the L1 amino terminus using the pGEX-2T vector (Pharmacia). GST-L1 fusion proteins were detected by SDS-PAGE of whole cell lysates (Figure 1, lanes 2-5), indicating accumulation to high levels.

Purification of HPV11 and HPV16 L1

When expressed at 37°C, the GST-L1 fusions were found in the insoluble cell pellet. When expression was induced at room temperature (rt), however, the GST-L1 fusions accumulated to high level, and partitioned into the soluble fraction of the cell lysates. When the soluble fraction was chromatographed on glutathione-Sepharose, the GST fusions were retained on the column, as shown by the 80 kDa species seen by SDS-PAGE in Figure 1, lanes 6 and 7. An additional intense 60 kDa band was present (Figure 1, lanes 6 and 7),
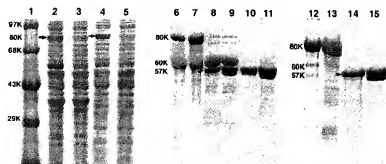

Figure 1. Purification of HPV11 and HPV16 L1 proteins after recombinant expression in *E. coli*. Shown is an SDS-PAGE analysis of GST-L1 and thrombin-released L1 proteins. Protein standards (lane 1); whole cell lysates induced and uninduced for HPV11 GST-L1 (lanes 2 and 3) and HPV16 GST-L1 (lanes 4 and 5) expression; GST-L1 fusions bound to glutathione resins for HPV11 (lane 6) and HPV16 (lane 7); glutathione affinity column eluates after thrombin cleavage for HPV11 (lane 8) and HPV16 (lane 9); L1 proteins precipitated after ATP-MgCl$_2$ treatment of the eluates in lane 8 (lane 10) and lane 9 (lane 11); glutathione affinity column eluates for HPV11 (lane 12) and HPV16 GST-L1 (lane 13) for preparations using the complete re-folding protocol (see Materials and Methods); eluates after thrombin cleavage from glutathione affinity column for HPV11 (lane 14) and HPV16 (lane 15) L1 proteins after using the re-folding purification protocol. The GST-L1 proteins of both virus types have an apparent molecular mass of 80 kDa (lanes 2, 4, 6, 7, 12, and 13). The 60 kDa species (lanes 6, 7, 8, and 9) is GroEL, as determined by N-terminal sequencing. Without using the re-folding protocol (lanes 6 to 9) the GroEL quantitatively co-eluted with L1. After using the refolding protocol, the GroEL contamination was almost eliminated (compare lanes 8 and 9 with lanes 14 and 15).

which could not be removed by extensive washing with various non-denaturing buffers. N-terminal sequencing identified the 60 kDa band as GroEL, an essential *E. coli* chaperone protein, which together with groES assists the folding of many proteins (Chen & Sigler, 1999; Weissman *et al.*, 1994, 1995). Electron microscopy (EM) analysis also detected GroEL double heptamers in the column eluate (data not shown; Chen *et al.*, 1994).

Thrombin digestion of the column-bound GST-L1 released L1 from the GST column, and this L1 is seen as a band migrating just below the 60 kDa GroEL band (Figure 1, lanes 8 and 9). GroEL also eluted from the column together with the cleaved L1 (Figure 1, lanes 8 and 9). The co-eluted GroEL could not be separated from L1 by gel-filtration or ion-exchange chromatography, suggesting a tight association. ATP-MgCl$_2$ releases GroEL from its substrate (Lissin *et al.*, 1990; Viitanen *et al.*, 1990). When ATP and MgCl$_2$ were added to the L1-GroEL eluate, a white precipitate formed. SDS-PAGE demonstrated that the precipitate was pure L1 (Figure 1, lanes 10 and 11). L1 could be separated from GroEL by treating the thrombin eluate with 2.5 M urea, but again all the L1 released from GroEL precipitated. These findings suggested that L1 was bound to GroEL in a such a way that its release led to an insoluble product.

If L1 was bound to GroEL in a "non-native" conformation, ATP treatment in the presence of GroES should facilitate release of "native" L1 protein. Therefore, the whole cell lysate, which also contains GroES, was treated with ATP and MgCl$_2$. GST-L1 was then purified by glutathione-Sepharose chromatography, yielding a GST-L1 fusion with reduced GroEL contamination (about 10-20 %). A similar reduction in the amount of GroEL was observed if the whole cell lysate was incubated with 3.5 M urea (see Materials and Methods). However, if the cell lysate was first treated with ATP and then with 3.5 M urea, the level of GroEL co-purifying with GST-L1 was reduced to approximately 1 % of the total protein (Figure 1, lanes 12 and 13), and the yield of L1 was increased two- threefold. The L1 released from the column by thrombin digestion was soluble and homogeneous (Figure 1, lanes 14 and 15) and could be concentrated to approximately 20 mg/ml^{-1}.

The purified L1 protein of HPV11 migrated on SDS-PAGE with an apparent molecular mass of 57 kDa (Figure 1, lane 14). In the case of HPV16 L1, however, there was a species of approximately 54 kDa in addition to the 57 kDa band (Figure 1, lane 15). The lower-to-upper band ratio increased with storage at 4°C, suggesting proteolysis at either the amino or carboxy terminus. N-terminal sequencing of these two bands demonstrated an intact HPV16 L1 amino terminus (data not shown), indicating proteolysis at the carboxy terminus.

The purification protocols for the HPV16 and HPV11 proteins are described in detail in Materials and Methods. The protocol is the same for both HPV11 and 16, except that the highest urea concentration for washing the glutathione-Sepharose column is 3.0 M for HPV11 L1, but only 2.3 M for HPV16 L1. A urea concentration greater than 2.3 M denatured and precipitated the HPV16 protein. A urea concentration of 3.5 M is optimal, however, for extraction of both HPV11 and 16 L1 from the total cell lysates. Using these conditions, approximately 3-5 mg of near homogeneous L1 was obtained from one liter of bacterial cell culture.

L1 Characterization

The purified L1 protein was analyzed by FPLC gel-filtration chromatography and negative-stain EM analysis (data not shown). Superdex-200 chromatography demonstrated a protein peak of about 300 kDa, consistent with pentamers. EM analysis showed typical pentamer "donuts", as previously seen for both L1 and VP1 purified after recombinant expression.

When L1 pentamers were digested with trypsin, the full-length 57 kDa L1 was reduced to a 42 kDa species on SDS-PAGE (Figure 2); the trypsin-digested pentamers appeared similar to untreated pentamers in the electron microscope (Li *et al.*, 1997). N-terminal sequencing of the tryptic 42 kDa products recovered from SDS-PAGE identified an
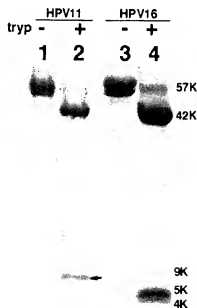

Figure 2. Trypsin digestion of purified HPV11 and HPV16 L1 proteins. Shown is an SDS-PAGE analysis of untreated (lanes 1, 3) and trypsin-digested HPV11 and HPV16 L1 proteins (lanes 2 and 4). In addition to a 42 kDa fragment common to both L1 types, trypsin digestion generated a 9 kDa product for HPV11 L1 and two species of approximately 4-5 kDa for HPV16 L1.

intact amino terminus for both for both HPV11 and HPV16 L1. N-terminal sequencing identified internal amino terminal sequences for HPV11 L1 as S_{348}-A-T-Y-T-N, and Y_{417}-V-Q-S-Q-A (corresponding to cleavage at K347 and K416), and for HPV16 L1 as N_{358}-T-N-F-K-E and F_{421}-V-T-S-Q-A (corresponding to cleavage at K357 and K420). A minor cleavage site was detected at K434 for HPV11 and K436 for HPV16.

When trypsin-digested L1 pentamers, presumably containing the 42 kDa fragments, were analyzed by Superdex-200 FPLC (15/60 column) they eluted at a position with an apparent molecular mass of 300 kDa, similar to that of the full-length L1. High-resolution SDS-PAGE analysis of the pentamer peak of trypsin-digested L1 from the Superdex-200 gel-filtration showed the 42 kDa band, and

an additional 9 kDa band for HPV11 (Figure 2, lane 2), and two other bands, of approximately 4 and 5 kDa for HPV16 (Figure 2, lane 4). Thus the smaller tryptic carboxy-terminal fragments remained associated with the larger 42 kDa fragment during gel-filtration chromatography, implying that the carboxy-terminal tryptic fragments are integral parts of the L1 pentamer structure.

Carboxy-terminal sequences required for L1 pentamer formation

To determine the sequences at the carboxy terminus necessary for proper L1 folding, we generated a series of carboxy-terminal deletions for both HPV11 and HPV16 L1. HPV11 L1 deletions were made as intein fusions to the L1 carboxy terminus,
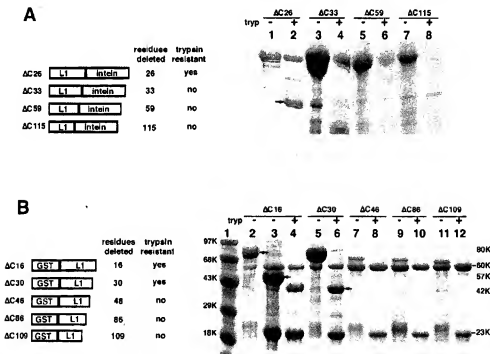
Figure 3. Pentamerization of carboxy-terminal deletion mutants of L1 proteins. The presence of a 42 kDa fragment after trypsin digestion was used as the criterion for proper folding and pentamer formation. (a) HPV11 L1 deletions were expressed as L1-intein fusions and whole cell lysates (incubated for 15 minutes with ATP-MgCl₂) were partially purified by chitin affinity chromatography. SDS-PAGE analysis shows the L1 proteins (cleaved from the L1-intein fusion by incubating the chitin resin with 20 mM DTT overnight followed by elution with three bed volumes of buffer L) before and after trypsin digestion. L1 was not well separated from GroEL (lanes 1, 3, 5, and 7). L1 degradation to smaller species was seen for ΔC33 and longer deletions. After trypsin digestion, only ΔC26 yielded the 42 kDa fragment (lanes 2, 4, 6, and 8). The bands remaining at the top of the gel are GroEL. (b) HPV16 L1 deletions were expressed as GST-L1 fusions and whole cell lysates (incubated for 15 minutes with ATP-MgCl₂) were partially purified by glutathione affinity chromatography. The protease was added directly to the GST-L1 fusions bound to glutathione-Sepharose. SDS-PAGE analysis shows the L1 proteins (as GST-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, the amounts of intact GST-L1 for ΔC46, ΔC86 and ΔC109 (lanes 7, 9, and 11) were decreased in comparison with ΔC16 and ΔC30 (lanes 2 and 5). Lane 3 was ΔC16 treated with thrombin. After trypsin digestion, although free GST was seen in all cases (lanes 4, 6, 8, 10, and 12), the 42 kDa product was seen only for ΔC16 and ΔC30. The 60 kDa GroEL remained unchanged. GST bands after trypsin digestion were more intense than the intact GST-L1 bands for ΔC46, ΔC86 and ΔC109, suggesting the presence of non-intact GST-L1 fusions.

while HPV16 L1 truncations were made as GST fusions to the L1 amino terminus (Figure 3). The criteria for correct pentamer oligomerization and folding was the presence of the 42 kDa trypsin-resistant species after trypsin digestion of the purified deleted proteins.

The HPV11 L1-intein constructs included ΔC26 (26 residues deleted), ΔC33, ΔC59 and ΔC115 (Figure 3(a)). The expressed fusion proteins of each construct were partially purified by chitin affinity chromatography (Figure 3(a)), and the column eluates containing both GroEL and L1 were digested with trypsin and analyzed by SDS–PAGE (Figure 3). Trypsin digestion of ΔC26 yielded the 42 kDa trypsin-resistant band on SDS–PAGE (Figure 3(a), lane 2), indicating proper folding. For the other three carboxy-terminal deletions, however, the 42 kDa band disappeared after trypsin digestion (Figure 3(a), lanes 3-8), indicating that the products of these deletions were in a non-native conformation. EM analysis of these proteins demonstrated the presence of pentamers only for the ΔC26 protein (data not shown), consistent with the trypsin digestion results.

Carboxy-terminal deletions of the HPV16 GST-L1 fusion behaved similarly to the HPV11 L1-intein fusions. The HPV16 GST-L1 deletions included ΔC16, ΔC30, ΔC46, ΔC86 and ΔC109 (Figure 3(b)). SDS–PAGE analysis of the glutathione-Sepharose before addition of trypsin showed the presence of both the GST-L1 deletion fusion proteins and GroEL (Figure 3(b), lanes 2, 5, 7, 9 and 11). Trypsin digestion of the ΔC16 and ΔC30 fusions yielded the 42 kDa band (Figure 3(b), lanes 4 and 6), ΔC46 and longer deletions of L1 were degraded completely (Figure 3(b), lanes 7-12). Again, EM analysis revealed that only ΔC16 and ΔC30 had a pentamer appearance. Only a single band was present on SDS–PAGE for the HPV16 ΔC30 protein after purification and thrombin cleavage (data not shown), corresponding to the lower band of the doublet for full-length L1 (Figure 1, lane 15). This result implies that a protease-sensitive site on full-length HPV16 L1 lies between the carboxy-terminal 26-30 residues, where there are indeed three lysine residues.

Amino-terminal sequences required for L1 pentamer formation

The N-terminal sequencing results showed that the L1 amino terminus is unaffected by trypsin digestion. In order to define the role of amino terminus region in L1 folding and pentamer formation, we made amino-terminal deletions in both HPV11 and HPV16 L1 (Figure 4). For L1-intein fusions of HPV11, ΔN5, ΔN8, ΔN14 and ΔN19 were expressed, the fusion proteins partially purified (Figure 4(a)) and subjected to trypsin digestion. The trypsin-resistant 42 kDa band was present for ΔN5 and ΔN8 (Figure 4(a), lanes 1, 2 and 5, 6), faintly present for ΔN14, and almost absent for ΔN19 (Figure 4(a), lanes 3, 4 and 7, 8). This result suggests that ΔN5 and ΔN8 proteins fold into trypsin-resistant pentamers, whereas longer deletions result in either only small amounts of the folded pentamers (ΔN14) or none at all (ΔN19).

For HPV16 GST-L1, the deletion constructs are shown in Figure 4(b). Trypsin digestion of the partially purified GST-L1 yielded the 42 kDa trypsin-resistant band for ΔN8, ΔN10 and to some extent for ΔN13 (Figure 4(b), lanes 1, 2, 3 and 5, 6, 7). The GST-L1 fusion of ΔN13 was not efficiently digested by trypsin (Figure 4(b), lanes 3 and 7); some full-length GST-L1 fusion was present even after prolonged incubation with trypsin. For ΔN20 (and ΔN30, not shown), the GST-L1 fusion proteins and GroEL were not detected by SDS–PAGE (Figure 4(b), lane 4). Instead, a broad band with the mobility of GST was present (Figure 4(b), lanes 4 and 8). It is possible that the L1 moieties of these GST-L1 fusions were degraded after expression but before purification.

Effect of deletions on in vitro capsid assembly

HPV16 L1 deleted proteins capable of pentamer formation were tested for in vitro self-assembly into higher-order structures (Table 1). All the amino-terminal and carboxy-terminal deletions except ΔN10 assembled into $T = 7$ particles under the conditions used (Figure 5(a)), but ΔN10

Table 1. The assembly of HPV16 L1 deletion mutants under different pH values

Constructs	Size of assembled particles				
	pH 4.0	pH 5.4	pH 6.2	pH 7.5	pH 8.5
Δ N-0	T7	T7	T7	T7/pent	pent/T7
Δ N-4	NT	T7	T7	T7/pent	NT
Δ N-6	NT	T7	T7	T7/pent	NT
Δ N-8	NT	T7	T7	T7/pent	pent/T7
Δ N-10	T1	T1	T1	T1	pent
Δ N-10 + Gly	T7	T7	T7	NT	NT
Δ N-10 + Thr(ΔN-9)	T7	T7	T7	NT	NT
Δ C-16	T7	T7	T7	NT	NT
Δ C-30	T7	T7	T7	NT	NT

Assembly properties of different HPV16 L1 truncation mutants at different pH. The assembly buffers were: pH 4.0 and 5.4, 40 mM sodium acetate; pH 6.2 and 7.5, 40 mM Hepes; pH 8.5, 40 mM Tris-HCl; all contained 1 M NaCl. NT, not tested.
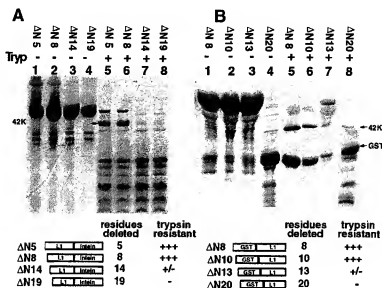
Figure 4. Pentamerization of amino-terminal deletion mutants of L1 proteins. The presence of a 42 kDa fragment after trypsin digestion was used as a criterion for proper folding and pentamer formation. (a) HPV11 L1 deletions were expressed as L1-intein fusions and whole cell lysates (incubated for 15 minutes with ATP-MgCl₂) were partially purified by chitin affinity chromatography. SDS-PAGE analysis shows the L1 proteins (as intein-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, full-length fusions are seen at the top of the gel for ΔN5 and ΔN8 (lanes 1 and 2). This band intensity decreases for ΔN14 (lane 3) and almost disappears for ΔN19 (lane 4). There is an intense band at the position of 60 kDa for each lane, which contains mostly intein-CBP (Chitin Binding Protein) and a small amount of CroEL (confirmed by N-terminal sequencing). The free intein-CBP implies extensive proteolysis of the fusion between L1 and intein. The cleaved CBP (together with some GroEL) does not bind chitin beads during purification. After trypsin treatment, the 42 kDa band is present for ΔN5 and ΔN8 (lanes 5 and 6), to some extent for ΔN14 (lane 7), but absent for ΔN19 (lane 8). Free chitin-CBP is digested to smaller fragments (lanes 5 and 8). (b) HPV16 L1 deletions were expressed as GST-L1 fusions, and whole cell lysates (incubated for 15 minutes with ATP-MgCl₂) were partially purified by glutathione affinity chromatography. SDS-PAGE shows the L1 proteins (as GST-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, full-length fusions for ΔN8, ΔN10, and ΔN13 migrate at the top of the gel (lanes 1, 2, and 3). Only free GST is present for ΔN20, possibly due to proteolytic degradation of the L1 portion during purification (lane 4). After trypsin digestion, the 42 kDa fragment is present for ΔN5 and ΔN10 (lanes 5 and 6). ΔN13 shows some 42 kDa and undigested fusion proteins (lane 7).

assembled into $T = 1$ particles under all conditions tested (Figure 5(b)). The precise deletion was critical. When even one Gly or Thr residue (the tenth L1 residue) was added to the amino terminus of ΔN10 (ΔN10 + Gly or ΔN10 + Thr), the proteins assembled into $T = 7$ particles (Figure 5, (c) and (d)).

The appearance of assembled products for each construct were generally the same in buffers with different pH (Table 1). The most consistent results were obtained at pH 5.4. Clumping of pentamers tended to become more severe at high pH, possibly due to disulfide bond formation of the free cysteine residues, and oligomeric pentamers (aggregates) were the predominant species at pH 8.5.

Recently, the ΔN10 deletion of HPV16 L1 was crystallized as a $T = 1$ particle, and its X-ray structure determined (Chen *et al.*, 2000). The structure showed that the carboxy terminal segment from residue 384-446 folds into three helices with connecting loops and turns (Figure 6). These helices

constitute the primary interpentamer bonding contacts in the assembled $T = 1$ particle. To test whether these helices also affect particle assembly, L1 proteins were generated with a specific deletion of helix 4 (see the structure in Figure 6) for both HPV16 (residues 408-431), and HPV11(residues 409-429). Pentamers were purified for these deleted proteins, as demonstrated by their FPLC elution profile and a donut appearance seen by EM. No assembly of particles from these pentamers was found under any condition tested (Figure 5(e)), suggesting that this carboxy-terminal helical domain is essential for particle assembly.

Discussion

Previous efforts to express HPV L1 in *E. coli* have produced only modest quantities of assembly-competent pentamers. The protocol described here gives a substantially enhanced yield of pure L1, using a relatively straightforward
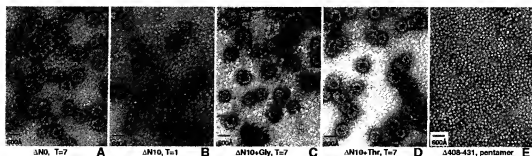
Figure 5. *In vitro* self-assembly of carboxy and amino-terminal deleted HPV16 L1 proteins. Electron micrographs of assembly reactions conducted in 1 M NaCl and 40 mM sodium acetate (pH 5.4), for purified L1 proteins. The bar length represents 600 Å, the diameter of a papillomavirus virion. (a) $T = 7$ particles (with diameter of about 600 Å) assembled from ΔN0 (full-length L1), with many unassembled pentamers in the background. (b) $T = 1$ particles (diameter of about 300 Å) assembled from ΔN10. (c) and (d) show the assembly products for ΔN10 + Gly and ΔN10 + Thr. The particles in (c) and (d) have a diameter of approximately 600 Å ($T = 7$), like those of ΔN0. (e) Free pentamers are the only products of *in vitro* assembly for Δ409-431L1.

purification strategy. Two factors appear to have been critical: (1) the use of GST-L1 fusion proteins, which greatly facilitated purification; and (2) a purification strategy that separated native L1 from the GroEL-bound complex. Despite the 27 kDa amino-terminal addition of the GST moiety, the fusion polypeptides fold and form pentamers properly, as evidenced by the release of L1 pentamers from the fusions by thrombin and their ability to be crystallized (Chen *et al.*, 2000).

GroEL release was accomplished by adding ATP-MgCl₂ to the cell lysate, followed by treatment with 3.5 M urea. We observed that 3.5 M

urea alone released about half of the GroEL. EM analysis demonstrated that GroEL double hepta-mers began to disassemble into single subunits in 2.0 M urea, while L1 pentamers withstood these urea concentrations. ATP-MgCl₂ alone also released about half of the GroEL. Neither urea nor ATP could be used used to release GroEL from the puri-fied L1-GroEL complex. Factors in the cell lysate presumably participated in the success of both treatments. In the case of ATP treatment, these fac-tors may have included GroES, which is required for normal ATP-dependent release of proteins from
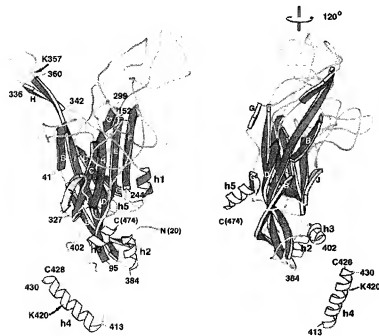
Figure 6. Three-dimensional structure of the HPV16 L1 monomer (Chen *et al.*, 2000). The orientation in (b) is rotated approximately 120° about the verti-cal axis from the orientation in (a). The β-strands are labeled by letter B through J. The starting residues of most β-strands and the starting and ending residue numbers for the carboxy-terminal α-helices are identified. Trypsin-sensitive sites are marked by arrows. Deletion of h4 (helix 4) of both HPV11 and HPV16 L1 abolished assembly of particles (see Figure 5 (e)).

GroEL, but why the urea treatment was unsuccessful when applied later in the purification is less evident.

How can we picture the GroEL/GST-L1 complex? The staining ratio in Figure 1 indicates about three GroEL subunits per GST-L1, or about one GroEL double heptamer per L1 pentamer. Since there is substantially less than one GroEL double heptamer per L1 monomer, the L1 must be largely folded and assembled, even with GroEL attached. The release of the GroEL by treatment with ATP-MgCl$_2$ in the presence of factors from the *E. coli* lysate may therefore proceed as a "normal" completion of the GroE cycle. Successful use of urea to dissociate GroEL and to produce assembly-competent L1 pentamers would require that the part of L1 interacting with GroEL be able to fold spontaneously once released.

The preparation of truncated forms of HPV11 and HPV16 L1 allowed further definition of the L1 domains required for pentamer and particle formation beyond that previously reported (Chen *et al.*, 2000). Up to 30 residues can be deleted from the carboxy terminus and at least ten residues from the amino terminus, without affecting pentamer formation. The amino-terminal 20 residues and the carboxy-terminal 30 residues are disordered in the crystal structure, and the deletion analysis thus shows that residues not part of the core of L1 (which we define as residues 21-474 for HPV16 and 19-471 for HPV11) are also not essential for pentamer assembly. A similar observation has been reported for polyomavirus VP1, where a protein with 32 residues deleted at the amino terminus and 63 residues deleted at the carboxy terminus can still form a pentamer amenable to crystallization and structure analysis (Stehle *et al.*, 1994). As in the case of L1, deletions compatible with pentamer assembly are ones that leave the protein core intact.

The carboxy-terminal sequences of L1 from different papillomaviruses are particularly variable. For example, the similarity of HPV16 and HPV11 L1 sequences ceases abruptly 30 residues from the carboxy terminus, just C-terminal to helix h5 (Chen *et al.*, 2000). Immediately preceding this boundary in the sequence are the residues that participate in the primary interpentameric contacts. Helices h2, h3, and h4 (see Figure 6) constitute a small subdomain that projects laterally from the β-jelly roll of L1, and helix h5 anchors this interaction domain by inserting back into the central cavity of the L1 pentamer. Data presented here, and related results in the literature, demonstrate that an intact h5 is essential for pentamer assembly as well as for subsequent capsid formation. For example, BPV L1 forms capsids with a 24 residue carboxy-terminal deletion, but not with a 40 residue deletion (Paintsil *et al.*, 1996); likewise, stable folding of canine oral papillomavirus (COPV) L1 tolerates a 26 residue, but not a 67 residue, deletion (Chen *et al.*, 1998). Our finding, that deletion of even three residues beyond the 30 residue boundary in

HPV11 L1 (HPV11-L1 ΔC33) produces an unfolded, trypsin-sensitive endproduct, demonstrates the importance of the complete L1 core for pentamer stability, including the helical subdomain and its h5 anchor. Our observation that the 42 kDa fragment obtained by limited proteolysis is not an independently stable entity leads to the same conclusion.

Because the helical subdomain of L1 is necessary for the integrity and stability of the pentamer itself, in addition to being essential for interactions among pentamers in the capsid, we presume that it adopts its folded structure prior to viral assembly. That is, a carboxy-terminal segment does not exit the L1 monomer to form a helix 5 contact with an L1 in a neighboring pentamer. In contrast, the entire, 63 residue, carboxy-terminal arm of polyomavirus VP1 is disordered on the subunits of free pentamers, and it folds into ordered structures only when it "invades" a neighboring pentamer as capsids assemble (Liddington *et al.*, 1991).

How does truncation of the amino terminus of L1 affect the design of the assembled particle? In a *T* = 1 small VLP, all of the carboxy-terminal interaction domains contact neighbors in identical, symmetrically related ways. In the 72 pentamer shells of virions, there must be some variability in the way these domains interact. Moreover, the curvature of the virion shell is much smaller than that of the small VLP, so the carboxy-terminal domains must face their interaction partners at different angles from that found in the *T* = 1 particles. The amino-terminal residues of the polypeptide chain lie immediately adjacent to the carboxy-terminal interaction domain, and it is reasonable to postulate that their presence prevents the high-curvature interaction seen in the small VLPs. A structure for the 72 pentamer shell would be required to understand precisely why the switch to small-shell assembly occurs so abruptly with deletion of the tenth amino-terminal residue.

In summary, the protocol described in detail in this paper permits expression and purification of large quantities of HPV L1 proteins. We have already used this procedure to determine the structure of HPV 16 L1, and we anticipate that the same methods will be useful for crystallizing L1 proteins from additional HPV types. The resulting pentamers and *in vitro* assembled particles will provide reagents for co-crystallization efforts, new substrates for immunologic studies, and potential new vaccine immunogens.

Materials and Methods

Cloning and deletion constructs

L1 clones of HPV 11 from human laryngeal (accession no. M14119) and of an HPV16 strain from human cervical carcinoma (accession no. AF140325) were used for this study. L1 PCR products were inserted downstream of GST coding sequence of pGEX-2T vector (Pharmacia) to generate GST fusions at the amino terminus of L1. For

HPV11 L1, primer 1 (forward) 5′ ACCTAC AGATCT
GGT ATGTGGCGTCCTAGCGACAGC and primer 2
(reversal) 5′ CA GAATCC TTACTTTTTGGTTTTGG-
TACGTT were used for amplifying the L1 coding
sequence and the PCR fragment were digested with
Bgl II/*EcoR* I for inserting into *BamHI/EcoRI*-digested
pGEX2T vector. The primers for HPV16 L1, forward 5′
CTGA AGATCT GGT ATG TCC CTG TGG CTG CCT
AGT GAG GCC AAT GTC and reversal 5′ CAG
GATATC TTAACGCTTACGTTTTTTGCGTTTAGC were
used for PCR and the PCR product was digested with
BglII/EcoRV for inserting into *BamHI/SmaI* sites of
pGEX2T.

The carboxy-terminal fusion constructs of HPV11 L1
with intein, were generated by cloning L1 PCR frag-
ments into the pCYB1 vector (IMPACT I system, NEB
Biolabs). The primers for PCR cloning were: forward
5′ CAATGTCA CATATG TGGCGTCCATCTGAT
AGCAC TGTATACGTCCACCTCCAAACCCTGTATCC
and reversal 5′ AGACTT GCTCTTCC GCAGTTTT
GGTTTTGGTACGTTTTCG. The PCR product was
digested with *NdeI/SapI* and then ligated into *NdeI/SapI*-
digested pCYB1 vector. *E. coli* strain XA90 [Δ*lacproXIII*,
nalA, argE(am), thi, rifr/F′lacIqZYproAB] was used for all
the cloning and expression studies. This strain consist-
ently expressed L1 fusions to higher level than other
strains tested.

Deletion mutants were constructed using PCR based
on the full-length L1 fusion clones. Sets of two primers
were used for PCR to amplify the whole plasmid con-
taining L1 gene except for the nucleotides to be deleted.
The PCR fragment contained a new unique restriction
site on both ends (incorporated into the primers) for
digestion and ligation. Clones were first identified by
digestion of the newly generated site and then further
confirmed by DNA sequencing. This strategy proved to
be a very efficient deletion method. Plasmids of 6-8 kb
were routinely amplified by PCR using the following
conditions: 94°C for 1.5 minutes, 65°C for 1 minute,
68°C 12 minutes, two cycles of PCR, then 94°C for 1.5
minutes, 68°C for 13 minutes, 23 cycles of PCR. The
reaction contained five units of Taq polymerase and 0.25
unit of Pfu (Stratagen). The inclusion of Pfu at the con-
centration of 1/20 Taq unit in the PCR reaction was
essential to obtain large quantities of full-length PCR
products for cloning.

Protein Purification

The detailed protocols of protein expression and puri-
fication of both HPV11 and HPV16 L1 are described as
follows. A 2 ml overnight cell culture was inoculated to
one liter of 2× YT at 37°C. Cells were grown untill
A_{600} to 0.1-0.3. The flasks were transferred to 25°C and
incubated with shaking for 30 minutes. IPTG was added
to 0.2 mM and protein expression induced for 16-20
hours. Cells from a one liter culture were resuspended in
30 ml of buffer L (50 mM Tris-HCl (pH 8.0), 0.2 M NaCl,
1 mM DTT, 1 mM EDTA) and lysed by sonication. After
sonication, ATP and MgCl$_2$ were added to final concen-
trations of 2 mM and 5 mM, respectively. After one hour
incubation at room temperature, urea (ultrapure grade)
was slowly added to the lysate to a final concentration
of 3.5 M. The mixture was incubated at room tempera-
ture for one hour with gentle shaking, and then dialyzed
against three changes of buffer over 18 hours. The lysate
was clarified by centrifugation at 25,000 g for 75 minutes.
The GST-L1 (or L1-intein) in the supernatant was

purified using a glutathione affinity column (or chitin
chromatography) using 15 ml bed volume for a lysate
from eight liters of bacteria. After washing the column
with ten bed volumes of buffer L, residual GroEL was
eluted by washing with ten bed volumes of 3.0 M urea
(in buffer L) for HPV11 L1 or 2.3 M urea for HPV16 L1.
L1 was cleaved from the GST fusion using thrombin
(Sigma, T6634) in an approximate ratio of 100 µg GST-L1
to one NIH unit of enzyme. The digestion was carried
out at 4°C overnight. L1 was further purified by Super-
dex-200 (16/60 column) gel-filtration FPLC. L1-intein
fusions were cleaved by incubation with DTT at 4°C
overnight (IMPACT I system of New England BioLabs).

Trypsin Digestion and *In vitro* Assembly

Limited trypsin digestion of L1 (trypsin to L1 ratio of
1:1000) was performed at room temperature, and the
digestion was terminated by adding 1 mM PMSF and an
equal volume of 2× SDS sample buffer followed by
immediate boiling. Assembly assays were conducted
using two methods. First, L1 protein was dialyzed (at a
concentration of approximately 0.1 mg/ml) against the
assembly buffer in a micro-dialysis button (Pierce) for 60
minutes. Second, concentrated L1 protein (10-20 mg/ml)
was diluted 100-1000-fold directly into assembly buffer
and incubated at room temperature for 30 minutes. Both
methods gave equivalent results. For EM analysis, the L1
protein was spotted on glow-discharged, carbon coated
grids (EM Science) and stained with 2% (w/v) uranyl
acetate. A JEOL 100-CX electron microscope was used
for visualization.

References

Baker, T. S., Newcomb, W. W., Olson, N. H., Cowsert,
L. M., Olson, C. & Brown, J. C. (1991). Structures of
bovine and human papillomaviruses. Analysis by
cryoelectron microscopy and three-dimensional
image reconstruction. *Biophys. J.* 60, 1445-1456.

Belnap, D. M., Olson, N. H., Cladel, N. M., Newcomb,
W. W., Brown, J. C., Kreider, J. W., Christensen,
N. D. & Baker, T. S. (1996). Conserved features in
papillomavirus and polyomavirus capsids. *J. Mol.
Biol.* 259, 249-263.

Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M.,
Jansen, A. M., Peto, J., Schiffman, M. H., Moreno,
V., Kurman, R. & Shah, K. V. (1995). Prevalence of
human papillomavirus in cervical cancer: a world-
wide perspective. International biological study on
cervical cancer (IBSCC) Study Group. *J. Natl Cancer
Inst.* 87, 796-802.

Chen, L. & Sigler, P. B. (1999). The crystal structure of a
GroEL/peptide complex: plasticity as a basis for
substrate diversity. *Cell*, 99, 757-768.

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P.,
Burston, S. G., Ranson, N. A., Clarke, A. R. &
Saibil, H. R. (1994). Location of a folding protein

and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature*, **371**, 261-264.

Chen, X., Garcea, R., Goldberg, I., Casini, G. & Harrison, S. C. (2000). Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol. Cell*, **5**, 557-567.

Chen, Y., Ghim, S. J., Jenson, A. B. & Schlegel, R. (1998). Mutant canine oral papillomavirus L1 capsid proteins which form virus-like particles but lack native conformational epitopes. *J. Gen. Virol.* **79**, 2137-2146.

Garcea, R. L., Salunke, D. M. & Caspar, D. L. (1987). Site-directed mutation affecting polyomavirus capsid self-assembly in vitro. *Nature*, **329**, 86-87.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D. R. & Schiller, J. T. (1992). Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proc. Natl Acad. Sci. USA*, **89**, 12180-12184.

Le Cann, P., Coursaget, P., Iochmann, S. & Touze, A. (1994). Self-assembly of human papillomavirus type 16 capsids by expression of the L1 protein in insect cells. *FEMS Microbiol. Letters*, **117**, 269-274.

Li, M., Cripe, T. P., Estes, P. A., Lyon, M. K., Rose, R. C. & Garcea, R. L. (1997). Expression of the human papillomavirus type 11 L1 capsid protein in *Escherichia coli*: characterization of protein domains involved in DNA binding and capsid assembly. *J. Virol.* **71**, 2988-2995.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). Structure of simian virus 40 at 3.8 Å resolution. *Nature*, **354**, 278-284.

Lissin, N. M., Venyaminov, S. & Girshovich, A. S. (1990). (Mg-ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature*, 348, 339-342.

Paintsil, J., Muller, M., Picken, M., Gissmann, L. & Zhou, J. (1996). Carboxyl terminus of bovine papillomavirus type-1 L1 protein is not required for capsid formation. *Virology*, 223, 238-44.

Pisani, P., Parkin, D. M. & Ferlay, J. (1993). Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. *Int. J. Cancer*, 55, 891-903.

Rayment, I., Baker, T. S., Caspar, D. L. & Murakami, W. T. (1982). Polyoma virus capsid structure at 22.5 Å resolution. *Nature*, 295, 110-115.

Rose, R. C., Bonnez, W., Reichman, R. C. & Garcea, R. L. (1993). Expression of human papillomavirus type 11 L1 protein in insect cells: *in vivo* and *in vitro* assembly of viruslike particles. *J. Virol.* 67, 1936-44.

Salunke, D. M., Caspar, D. L. & Garcea, R. L. (1989). Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* 56, 887-900.

Stehle, T., Yan, Y., Benjamin, T. L. & Harrison, S. C. (1994). Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature*, 369, 160-163.

Trus, B. L., Roden, R. B., Greenstone, H. L., Vrhel, M., Schiller, J. T. & Booy, F. P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution. *Nature Struct. Biol.* 4, 413-420.

Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P. & Lorimer, G. H. (1990). Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K+ dependent. *Biochemistry*, 29, 5665-5671.

Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell*, 78, 693-702.

Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. & Horwich, A. L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell*, 83, 577-587.

Xi, S. Z. & Banks, L. M. (1991). Baculovirus expression of the human papillomavirus type 16 capsid proteins: detection of L1-L2 protein complexes. *J. Gen. Virol.* 72, 2981-2988.

Zhou, J., Doorbar, J., Sun, X. Y., Crawford, L. V., McLean, C. S. & Frazer, I. H. (1991). Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. *Virology*, 185, 625-232.

Expression of the Human Papillomavirus Type 11 L1 Capsid Protein in *Escherichia coli*: Characterization of Protein Domains Involved in DNA Binding and Capsid Assembly

MAOLIN LI,[1] TIMOTHY P. CRIPE,[1]† PATRICIA A. ESTES,[1] MARY K. LYON,[2]
ROBERT C. ROSE,[3] AND ROBERT L. GARCEA[1]*

Section of Pediatric Hematology/Oncology, Department of Pediatrics, University of Colorado School of Medicine, Denver, Colorado 80262[1]; Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309[2]; and Infectious Disease Unit, Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642[3]

The L1 major capsid protein of human papillomavirus type 11 (HPV-11) was expressed in *Escherichia coli*, and the soluble recombinant protein was purified to near homogeneity. The recombinant L1 protein bound DNA as determined by the Southwestern assay method, and recombinant mutant L1 proteins localized the DNA-binding domain to the carboxy-terminal 11 amino acids of L1. Trypsin digestion of the full-length L1 protein yielded a discrete 42-kDa product (trpL1), determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, resulting from cleavage at R415, 86 amino acids from the L1 carboxy terminus. Sucrose gradient sedimentation analysis demonstrated that trpL1 sedimented at 11S, while L1 proteins with amino-terminal deletions of 29 and 61 residues sedimented at 44S. Electron microscopy showed that the full-length L1 protein appeared as pentameric capsomeres which self-assembled into capsid-like particles. The trpL1 protein also had a pentameric morphology but was unable to assemble further. In an enzyme-linked immunosorbent assay, the trpL1 and L1 capsids reacted indistinguishably from virus-like particles purified after expression of HPV-11 L1 in insect cells. The carboxy terminus of L1 therefore constitutes the interpentamer linker arm responsible for HPV-11 capsid formation, much like the carboxy-terminal domain of the polyomavirus VP1 protein. The trypsin susceptibility of HPV-11 L1 capsids suggests a possible mechanism for virion disassembly.

Papillomaviruses are a family of nonenveloped, double-stranded DNA viruses that infect many species, and at this time more than 70 different types of human papillomaviruses (HPVs) have been identified. Image analysis of cryoelectron micrographs of bovine papillomavirus type 1 (BPV-1) and HPV-1 has demonstrated a common structure comprised of 72 pentameric capsomeres arranged on a T=7 icosahedral capsid lattice (2). This overall structure is similar to that previously described for simian virus 40 (SV40) and murine polyomavirus (15, 27), although possible differences in capsomere morphology and intercapsomere contacts are suggested by the image reconstruction. The papillomavirus genome encodes two structural proteins, L1 and L2, synthesized late in infection. In the virion, the ratio of L1 (55 kDa) to L2 (74 kDa) has been estimated over a range from 5:1 to 30:1 (26a). By analogy to polyomavirus and SV40, the capsomeres are likely formed by pentamers of L1. L2 may associate with the L1 pentamer, although given the uncertain ratio of L1 to L2, all or a subset of the pentamers may be occupied.

Papillomaviruses are difficult to propagate in vitro. Therefore, recombinant expression systems have been used to express the L1 and L2 proteins in order to characterize their immunologic characteristics as well as to analyze their structural properties. Capsid proteins for HPV-16, BPV-1, HPV-11, and HPV-33 have been expressed in insect cells with baculovirus vectors (12, 23, 28), and those for HPV-11, HPV-16, and BPV-1 have been expressed in mammalian cells with vaccinia virus (10, 30). These expression systems all generate virus-like particles (VLPs) similar in appearance to empty capsids, and these VLPs have immunologic characteristics suggesting a native conformation. When cells expressing L1 are examined by electron microscopy, the VLPs are seen exclusively in the nuclei of the cells (12, 23, 30). These results are also analogous to observations with polyomavirus (17), suggesting not only common structural features between these virus families but also similar control of the intracellular site of capsid assembly.

The polyomavirus major capsid protein VP1, purified after expression in *Escherichia coli*, self-assembles in vitro into capsid-like structures (25). This self-assembly reaction has provided a useful method for assaying conditions favoring capsid assembly as well as for examining mutant VP1 proteins for their ability to assemble (8, 26). In order to analyze the assembly properties of L1, we have purified the HPV-11 L1 protein after expression in *E. coli*. This recombinant protein also self-assembled in vitro into capsid-like particles; however, truncations of the L1 protein revealed a domain structure that differs from that of the polyomavirus VP1 protein.

MATERIALS AND METHODS

---

* Corresponding author. Mailing address: Section of Pediatric Hematology/Oncology, Department of Pediatrics, University of Colorado School of Medicine, 4200 E. Ninth Ave., Box C229, Denver, CO 80262.
† Present address: University of Wisconsin Comprehensive Cancer Center, Madison, WI 53792.

The mutant L1 coding sequences with amino-terminal and carboxy-terminal deletions were generated with the cloning methods described above. Two proteins with amino-terminal deletions, NΔ29L1 and NΔ61L1, were PCR amplified with two forward primers, 5'-GCGAAGCTTCATATGACCAACATATTTTAT CATGCC and 5'-GCGAAGCTTCATATGGTGTCTGGATATCAATATAGA, and a reverse primer as described above. Two mutant L1 proteins with carboxy-terminal deletions, CΔ11L1 and CΔ86L1, were PCR amplified with a forward primer as described above and two reverse primers, 5'-CCTGGATCCAGATC TCATGTAGAGGGCTTAGACACAGC and 5'-CCTGGATCCAGATCTCAT CTATAAGTATCCTCGAGTGT.

Recombinant L1 expression and purification. L1 proteins were expressed from the T7 promoter of pET-3a-HPV-11 L1 in the host bacteria BL21(DE3). Bacteria were grown at 37°C in M9 medium with ampicillin (100 μg/ml) overnight. Fresh medium (50 ml) was then inoculated with 1 ml of this overnight culture, and cells were grown for 6 to 8 h at 37°C. This culture was used to inoculate 500 ml of fresh medium containing 100 μg of ampicillin per ml. Bacteria were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for 6 h. The cells were then pelleted and stored at −20°C.

Bacteria from 2 liters of culture were thawed in 200 ml of buffer A with 0.25 M NaCl (50 mM Tris-HCl [pH 7.9], 5% glycerol, 2 mM EDTA, 15 mM 2-mercaptoethanol, 250 mM NaCl), and lysozyme was added to 200 μg/ml. After 20 min at 4°C, sodium deoxycholate and phenylmethylsulfonyl fluoride were added to final concentrations of 0.05% (wt/vol) and 1 mM, respectively. All subsequent steps were carried out at 4°C. The cell suspension was sonicated with a Heat System Ultrasonics W-225 for three 40-s bursts, at 1-min intervals. The sonicated suspension was subjected to Dounce homogenization 20 times with a 8 pestle. The homogenate was centrifuged at 12,000 × g for 20 min, and Polymin P was added slowly to the supernatant to a final concentration of 0.5% (wt/vol) and incubated for 20 min. After centrifugation at 10,000 × g for 15 min, the pellet was resuspended in 50 ml of buffer B with 1 M NaCl (10 mM Tris-HCl [pH 7.9], 5% glycerol, 2 mM EDTA, 15 mM 2-mercaptoethanol, 1 M NaCl) with a Dounce homogenizer. The suspension was centrifuged at 10,000 × g for 15 min. The pellet was extracted with 50 ml of the same buffer, and the supernatants were combined and precipitated by the addition of ammonium sulfate to 35% saturation. The pellet from the first homogenate was extracted with 100 ml of buffer B with 1 M NaCl with a Dounce homogenizer. The suspension was centrifuged at 10,000 × g for 15 min, and the supernatant was precipitated with ammonium sulfate as described above. This precipitate was resuspended in 20 ml of buffer C with 0.1 M NaCl (10 mM Tris-HCl [pH 7.2], 5% glycerol, 2 mM EDTA, 15 mM 2-mercaptoethanol, 100 mM NaCl) and dialyzed against the same buffer. The insoluble material was removed by centrifugation at 10,000 × g for 20 min. The supernatant was subjected to chromatography on a DE-52 cellulose column in buffer C with 0.1 M NaCl. The flowthrough was applied to a P11 phosphocellulose (Whatman) column. The column was washed once with buffer C with 0.25 M NaCl and once with buffer C with 0.5 M NaCl. L1 was eluted with buffer C with 1 M NaCl. Further purification was carried out by pelleting the protein from sucrose gradients (see below) or directly from the 1 M NaCl eluant by centrifugation at 120,000 × g for 20 min. Purification steps were monitored by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie blue staining and immunoblotting (23).

Trypsin digestion and protein sequence determination. Purified L1 protein was dialyzed against trypsin digestion buffer (10 mM Tris-HCl, 2 mM EDTA, 15 mM 2-mercaptoethanol, 100 mM NaCl, 100 mM NaHCO₃, pH 7.8). Trypsin digestion was carried out for 2 h at 4, 25, or 37°C, with the ratio of protein to trypsin at 1:100 (wt/w). Trypsin-digested L1 peptides were resolved by SDS-7.5% PAGE and electroblotted to an Immobilon-P membrane in CAPS buffer (10 mM 3-[cyclohexylamino]-1-propanesulfonic acid, 10% methanol, pH 11.0) (1). Proteins were stained with 0.1% Coomassie blue R-250 in 50% methanol for 5 min and destained with 50% methanol–10% acetic acid with several changes. The L1 proteolytic fragments were excised and subjected to protein sequencing. The amino-terminal 8 amino acids of the trypsin products were determined with an Applied Biosystems gas phase sequencer.

Southwestern assays. L1 was separated by SDS–9% PAGE, transferred to nitrocellulose membranes, and then incubated for 20 min with 2% bovine serum albumin. Proteins were renatured in buffer D (50 mM Tris-HCl [pH 7.4], 1 mM EDTA, 200 mM NaCl, 10% glycerol, 0.1% Nonidet P-40) at 4°C overnight. The DNA probe was prepared from the plasmid SP65 digested with EcoRI and labeled with [³²P]dCTP, using the Prime-a-gene System (Promega). The DNA binding assay was carried out in buffer E (30 mM HEPES [pH 7.4], 5 mM MgCl₂, 50 mM NaCl) with ³²P-labeled plasmid DNA for 30 min at room temperature (19). The nitrocellulose filters were washed four times with binding buffer and then subjected to autoradiography.

Sucrose gradient sedimentation of L1 proteins. L1 fractions eluting from the phosphocellulose column in buffer C with 1 M NaCl were dialyzed against buffer A with 0.25 M NaCl. Four hundred microliters was layered onto a 5 to 20% sucrose gradient in buffer A with 0.25 M NaCl and centrifuged at 39,000 rpm for 16 h in an SW41 rotor. Fractions (0.5 ml) were collected from below. The fractions were assayed for L1 by immunoblotting or Southwestern assay following SDS-PAGE. The L1 protein pelleted in the sucrose gradients was not homogeneous.

Electron microscopy. L1 protein obtained after the phosphocellulose purification step was dialyzed into buffer containing 0.1 M NaCl, 20 mM Tris-HCl (pH

7.2), 1 mM EDTA, and 15 mM 2-mercaptoethanol. Samples were applied to glow-discharged carbon-coated 400-mesh grids and stained with 2% uranyl acetate. Images were photographed with a Philips CM 10 electron microscope at nominal magnifications of ×39,000 or ×73,000.

Enzyme-linked immunosorbent assay (ELISA). VLPs were purified as described previously (22, 23). Methods used for the production of rabbit polyclonal HPV-11 virion and HPV-11, HPV-16, and HPV-18 VLP antisera have been described previously (4, 22, 24). Purified antigens were diluted in phosphate-buffered saline to a concentration of 0.01 mg/ml, and 100-μl aliquots were dispensed into 96-well microtiter plates. The plates were incubated overnight at 4°C and then blocked for 2 h at room temperature, as previously described (24). Threefold serial dilutions of primary antisera were made in a diluent solution containing 1% bovine serum albumin (Kirkegaard & Perry Laboratories, Inc., Gaithersburg, Md.) and 10% (vol/vol) wild-type *Autographa californica* nuclear polyhedrosis virus-infected cell culture supernatant, added to reduce background seroreactivities (22). Diluted sera were preabsorbed overnight at 4°C, centrifuged for 1 min at 14,000 × g, and then added in 100-μl aliquots to wells containing antigen (test wells) or phosphate-buffered saline alone (control wells). Secondary antibody conjugate (anti-rabbit immunoglobulin G-alkaline phosphatase at a 1:2,000 dilution; Kirkegaard & Perry Laboratories) and enzyme substrate were then added as previously described (24). Specific absorbance values were calculated by subtracting control well values from test well values, and replicate differences were averaged.

RESULTS

Purification of the HPV-11 L1 protein after expression in E. coli. The full-length HPV-11 L1 coding sequence was cloned into the vector pET-3a to yield pET3aL1, and protein expression was induced from the T7 promoter of the vector by addition of IPTG. L1 was expressed more efficiently in the *E. coli* strain BL21(DE3) than in HMS174(DE3) (data not shown). The L1 protein was purified according to a protocol based on that used for the purification of the polyomavirus VP1 protein (14). The steps in the purification were assayed by SDS-PAGE and immunoblot analysis with an anti-L1 rabbit antiserum (Fig. 1) (23). After cell lysis, approximately 50% of the L1 was soluble. L1 remaining in the bacterial cell pellet at this stage could be partially recovered by further extraction in a buffer containing 1 M NaCl. Under the loading buffer conditions, L1 was found in the flowthrough fraction from the DE-52 cellulose column (Fig. 1, lanes 3) and was retained on the P11 phosphocellulose column (Fig. 1, lanes 4 and 5). L1 eluted in 1 M NaCl buffer from the phosphocellulose column (Fig. 1, lanes 6). This chromatographic behavior is similar to that of the polyomavirus VP1 protein (14). The eluant from the phosphocellulose column contained various amounts of minor contaminating proteins, and final purification was achieved with sedimentation through a sucrose gradient (Fig. 1, lanes 6 and 7).

L1 binds DNA. Analysis of L1 DNA binding was performed with the Southwestern blot assay (19). In order to characterize the DNA-binding domain identified by this assay, L1 proteins with amino- and carboxy-terminal deletions were generated by site-directed mutagenesis of the L1 expression vector. Mutant proteins were partially purified after expression in bacteria. The L1 proteins were separated by SDS-PAGE, electrotransferred to nitrocellulose, and hybridized with ³²P-labeled plasmid DNA. As seen in Fig. 2, DNA binding similar to that of the full-length protein was seen for two mutant proteins with amino-terminal deletions, NΔ29L1 and NΔ61L1. However, two mutant proteins with carboxy-terminal deletions, CΔ86L1 and CΔ11L1, failed to bind DNA. DNA binding to the full-length L1 protein was also tested with increasing concentrations of NaCl in the Southwestern blot assay, and DNA binding was lost when greater than 200 mM NaCl was present in the wash buffer (data not shown).

The mutant CΔ11L1 also behaved differently from the full-length L1 protein during its purification, in that CΔ11L1 eluted from the phosphocellulose column in a buffer containing 0.5 M
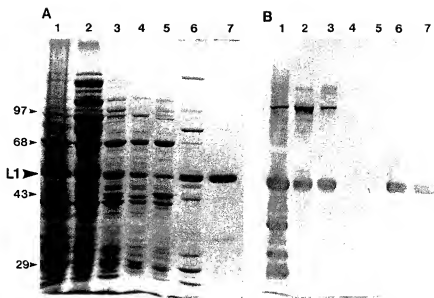
FIG. 1. Purification of HPV-11 L1 after expression in *E. coli*. SDS-PAGE analysis of L1 fractions either stained with Coomassie blue (A) or immunoblotted with an anti-L1 antiserum (B). Lanes: 1, whole-cell lysate; 2, ammonium sulfate precipitate; 3, flowthrough from DE-52 column; 4, flowthrough from phosphocellulose column; 5, 250 mM NaCl eluant from phosphocellulose; 6, 1 M NaCl eluant from phosphocellulose; 7, L1 pelleted after sucrose gradient sedimentation. Molecular mass standards (in kilodaltons) appear on the left.

NaCl. For the polyomavirus VP1 protein, disulfide reducing agents eliminated DNA binding in the Southwestern assay (19). However, neither 2-mercaptoethanol nor dithiothreitol affected L1 DNA binding (data not shown). A DNA-binding domain therefore can be localized to the carboxy-terminal 11 amino acids of L1.

Trypsinization of recombinant L1 yields a discrete 42-kDa fragment. Protease sensitivity of viral capsid proteins has been
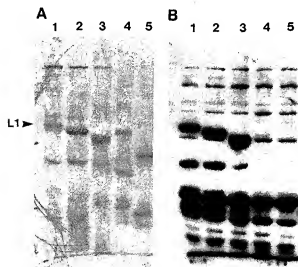

FIG. 2. Southwestern analysis of L1 mutant proteins. Recombinant L1 proteins were partially purified after expression in *E. coli*, subjected to SDS-PAGE, and either immunoblotted with an anti-L1 antiserum (A) or subjected to Southwestern analysis (B). Lanes: 1, wild type; 2, NΔ29L1; 3, NΔ61L1; 4, CΔ11L1; 5, CΔ86L1.

informative in assessing their native conformation, and for example, trypsin cleaves the polyomavirus VP1 protein at a single lysine residue (K28) near the amino terminus (19). In order to characterize protease-accessible domains of L1, the purified recombinant L1 protein was subjected to trypsin digestion. After digestion with trypsin, two progressive cleavage products were identified by SDS-PAGE (Fig. 3A). The first (L1a) was an ≈48-kDa species obtained after digestion at either 4 or 25°C. The second was an ≈42-kDa species (trpL1) which was obtained after digestion at 37°C for 2 h (Fig. 3A, lane 4) and which remained unchanged upon longer incubations. These cleavage products were characterized by amino-terminal protein sequencing after electrotransfer to an Immobilon filter. Both products had an intact amino terminus, indicating cleavage sites at the carboxy terminus. When directly compared, CΔ86L1 (Fig. 2A, lane 5) (calculated molecular mass, 45.7 kDa) migrated at the same apparent molecular weight by SDS-PAGE (Fig. 3A, lane 4), suggesting that R415 is the cleavage site used to generate trpL1. Based upon its SDS-PAGE mobility in relation to the full-length and trpL1 proteins, the L1a cleavage site is likely to be at residue K450, K460, or R471. The resistance of the amino terminus to digestion suggests that this domain of L1 is relatively inaccessible to digestion in comparison to the carboxy terminus.

In order to confirm the results of the previous DNA binding experiments, the trypsin-digested L1 cleavage products were analyzed with the Southwestern assay (Fig. 3C). Consistent with the CΔ86L1 and CΔ11L1 results, both trypsin cleavage products did not bind DNA in this assay. In addition, both trpL1 and L1a eluted from phosphocellulose in buffers containing 100 mM NaCl (data not shown), again correlating decreased phosphocellulose binding with deletion of the putative L1 DNA-binding domain.

Sedimentation analysis identifies inter- and intrapentamer bonding domains. Sedimentation analysis has been used to separate polyomavirus subviral intermediates as well as to as-
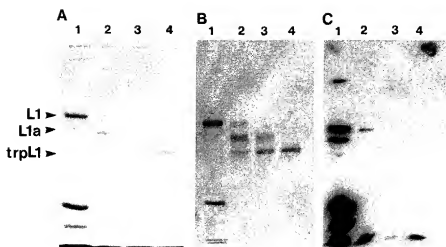
FIG. 3. Trypsin digestion of L1. L1 protein was partially purified after expression in *E. coli*, digested for 2 h with trypsin at the indicated temperatures; subjected to SDS-PAGE, and stained with Coomassie blue (A), immunoblotted with an anti-L1 antiserum (B), or subjected to Southwestern analysis (C). Lanes: 1, no trypsin at 37°C; 2 to 4, trpsin at 4, 20, and 37°C, respectively. L1a and trpL1 indicate major proteolytic products.

sess capsomere and capsid formation (9). For polyomavirus, the pentameric VP1 capsomere sediments at 7.5S, and the capsid sediments at 140S (25). To assay the effect of L1 mutations on capsomere and capsid formation, velocity sedimentation analysis of full-length, trypsinized, and mutant L1 proteins was performed. Given the molecular weight of L1, pentamers of ≈270 kDa might be expected to sediment at 10 to 14S. Under the conditions shown in Fig. 4, capsid-like particles and aggregates having a sedimentation value greater than 19S were pelleted to the bottom of the gradients. The associated nature of the full-length L1 protein was made evident by the small amount of protein sedimenting between 7 and 14S compared with the amount present in the pellet fraction. The

majority of the CΔ11L1 protein also pelleted, suggesting capsid or aggregate formation in this preparation (data not shown). The trpL1 protein showed a major component with a sedimentation coefficient of 11 to 12S, a value consistent with that expected for unassociated pentamers.

The mutant proteins with amino-terminal deletions, NΔ29L1 and NΔ61L1, precipitated during the dialysis before the DE-52 column, and only a small amount of soluble mutant protein was obtained. Therefore, their presence in the sucrose gradients was monitored by SDS-PAGE and with the Southwestern assay. Both proteins sedimented with a predominant species at 4S (Fig. 4) (data for NΔ61L1 are not shown). Furthermore, a Centricon-100 spin column (Amicon) with a 100-
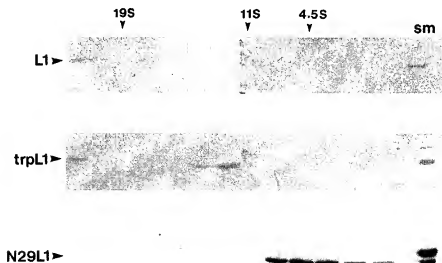

FIG. 4. Sucrose gradient sedimentation of L1 proteins. SDS-PAGE analysis of fractions after sedimentation of L1 proteins in 5 to 20% sucrose gradients. Proteins were detected by immunoblotting (L1 and trpL1) or Southwestern assay (NΔ29L1). Sedimentation markers in a parallel gradient were β-galactosidase (19S), catalase (11.3S), and hemoglobin (4.5S). sm, starting material.

kDa exclusion limit was used to concentrate the mutant NΔ29L1 and NΔ61L1 proteins. Under these conditions, pentameric capsomeres should be retained. Immunoblot analysis showed that the mutant proteins passed through the Centri-con-100 (data not shown), providing additional evidence that these truncations at the amino terminus affected pentamer formation and resulted in an L1 protein species of approximately monomer-to-dimer size.

Recombinant L1 self-assembles in vitro into capsid-like particles. HPV-11 L1 protein purified after recombinant baculovirus expression in Sf9 insect cells self-assembles into VLPs (23). Based upon this result, L1 protein purified after expression in *E. coli* was analyzed for in vitro capsid assembly by electron microscopy. As shown in Fig. 5A, the L1 protein, which eluted from the phosphocellulose column in 1 M NaCl, self-assembled into capsid-like structures with a diameter similar to that of papillomavirus virions (55 nm). In addition, unassembled capsomeres with a central stain-filled hollow, which appeared similar to the pentameric capsomeres of polyomavirus VP1, could also be identified (25). Electron micrographs of the trypsin digestion products, L1a and trpL1, showed a uniform population of pentamer-like structures (Fig. 5B and C), although many of the pentamers appeared to lie on their sides, presenting an H or V shape. Capsid-like structures were no longer present after trypsin digestion of L1. Higher-magnification views (Fig. 5C) verified the similarity of these structures to truncated pentamers of the polyomavirus VP1 protein (8). The CA11L1 protein was further purified by sucrose gradient sedimentation in order to obtain a homogeneous preparation suitable for electron microscopy. Electron micrographs demonstrated that CA11L1 is also capable of in vitro self-assembly into capsid-like aggregates, although they are of less uniform appearance than those of the full-length L1 protein (Fig. 5D). Because of the difficulty in purifying the CΔ86L1 protein, this mutant protein could not be analyzed by electron microscopy.

In order to test whether L1 assembly occurred in vivo, bacteria expressing L1 were fixed, sectioned, and directly examined by electron microscopy. No capsid-like particles were seen in the bacteria, although inclusion bodies were present (data not shown). In addition, crude bacterial lysates were subjected to sucrose gradient sedimentation, and fractions were assayed both by immunoblotting for L1 and by electron microscopy. Only rare capsid-like particles could be detected in the presence of abundant L1 protein (data not shown). Thus, capsid assembly occurs after bacterial lysis and is likely facilitated by the exposure of the L1 protein to high ionic strength, e.g., the phosphocellulose column elution buffer, or to other in vitro conditions which may favor capsid formation.

Immunologic reactivity of bacterially expressed HPV-11 L1. Results from previous studies have demonstrated that VLPs contain conformationally dependent capsid-neutralizing antigenic domains similar to those of native HPV-11 virions (24). In addition, these epitopes are antigenically distinct from similar epitopes of HPV-16 and HPV-18 (22). The antigenic properties of HPV-11 capsids and isolated capsomeres obtained after expression in *E. coli* were compared with VLPs in an ELISA (Fig. 6, top panel) (4, 24). Antisera raised against either native HPV-11 virions (4) or HPV-11 VLPs (24) reacted indistinguishably with all of the HPV-11 L1 antigens tested. Denatured VLPs reacted much less efficiently in this assay (data not shown), indicating that nonlinear epitopes were the predominant epitopes recognized by these antisera. Virus type specificity of the epitopes was verified by the failure of antibodies raised against either HPV-16 or HPV-18 capsids to react with the HPV-11 antigens (Fig. 6, bottom panel). These findings are

consistent with previous results (22) and indicate that HPV-11 capsids and L1 capsomeres purified after expression in *E. coli* are antigenically similar to HPV-11 VLPs and, thus, to native HPV-11 virions.

DISCUSSION

We have expressed and purified the HPV-11 L1 protein in *E. coli*. The purified protein is soluble and has many of the biochemical properties found previously for the recombinant polyomavirus VP1 protein, such as DNA binding, pentameric morphology, and the ability to self-assemble into capsid-like particles. The introduction of site-directed mutations into the expression vector has allowed identification of three functional domains of the L1 protein: (i) a DNA-binding region near the carboxy terminus, (ii) a region within 86 amino acids of the carboxy terminus required for intercapsomer bonding, and (iii) a region near the amino terminus which affects pentamer formation. The positions of these domains distinguish HPV-11 L1 from polyomavirus VP1, in that the VP1 DNA-binding domain is located at the amino terminus and the intercapsomere bonding domain is at the carboxy terminus. This difference suggests an alternative bonding strategy for L1 in the papillomavirus capsid compared to that of VP1 in the polyomavirus capsid.

Digestion with trypsin was used to identify accessible residues and thus provide evidence for conformationally exposed regions of the L1 pentamer. Characterizations of papillomavirus structural proteins derived from purified virions by SDS-PAGE have previously noted minor capsid protein species which may be degradation products related to those generated by trypsin digestion (6, 7). Interestingly, trypsin cleavage sites were found only at the carboxy terminus, extending inward approximately 86 residues. The inability of the resulting trpL1 pentamers to self-assemble while maintaining their pentameric morphology is similar to the behavior of the polyomavirus ΔNCO VP1 mutant protein (8). This result provides evidence that the intercapsomere linker arm of L1 is located at its carboxy terminus. The PPP motif around residue 404 (Fig. 6) may be similar to the PYP motif in polyomavirus (residue 320) and SV40 (residue 300) (15). These residues signal the exit of the polypeptide chain from the body of the pentamer and its redirection towards its capsomere neighbor, and therefore the PPP motif may represent the beginning of an interpentamer linker arm for L1. The ability of the CA11L1 mutant protein to assemble may therefore define the linker arm as the domain between L1 residues 407 and 490.

The sedimentation characteristics of the mutant and trypsinized L1 proteins were consistent with the morphology of the proteins determined by electron microscopy. The trpL1 protein is morphologically similar to the polyomavirus VP1 pentameric capsomere, and the 12S value is therefore representative of an L1 capsomere. The proteins with amino-terminal deletions, NΔ29L1 and NΔ61L1, sediment at less than capsomere size, suggesting that pentamer oligomerization is disrupted by these deletions. The small PPP motif, residues 12 to 14 (Fig. 7A), may indicate a chain redirection at the amino terminus of L1, and it is possible that deletions before this signal may not affect pentamer formation.

With the Southwestern assay, a DNA-binding domain was localized to the carboxy-terminal 11 residues of HPV-11 L1 (Fig. 2). We assume that this DNA binding is sequence nonspecific, because a plasmid probe was used for its identification and because similar results for the polyomavirus VP1 protein have indicated no preference for the source of DNA (19). A previous reported study of the HPV-11 L1 protein in which our
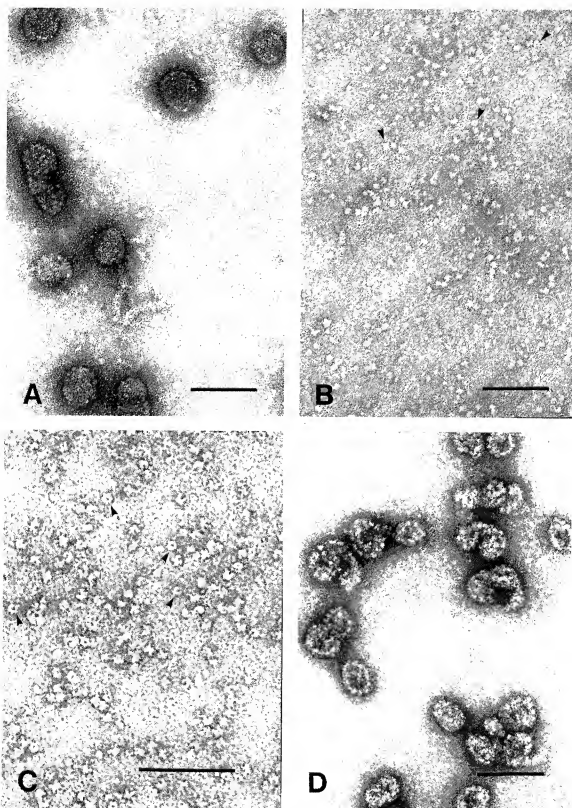
FIG. 5. Electron micrographs of L1 proteins after in vitro self-assembly. Shown are the full-length L1 protein (A), L1a (B), trpL1 (C), and CΔ11L1 (D). (B and C) Arrowheads indicate structures resembling the pentameric capsomeres of polyomavirus VP1 (8), with their stain-filled fivefold axis orthogonal to the grid. Bar, 100 nm.
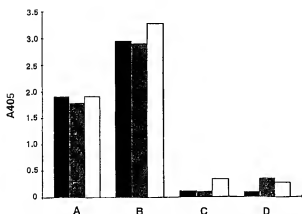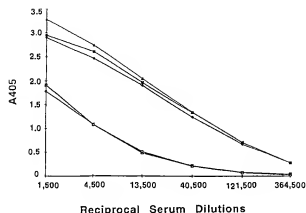
FIG. 6. Antigenic properties of the HPV-11 L1 protein purified after expression in *E. coli*. (Top panel) Polyclonal rabbit antisera against HPV-11 VLPs (filled symbols) or HPV-11 virions (open symbols) were compared in an ELISA for reactivity against VLPs (triangles), capsids prepared from L1 purified after expression in *E. coli* (squares), or trpL1 capsomeres (diamonds). (Bottom panel) Type-specific polyclonal HPV capsid antisera were tested for immunoreactivity in an ELISA against VLPs (white bars), L1 capsomeres (grey bars), or capsids assembled from L1 purified after expression in *E. coli* (black bars). Tested were HPV-11 virion antiserum (A), HPV-11 VLP antiserum (B), HPV-16 VLP antiserum (C), and HPV-18 VLP antiserum (D).
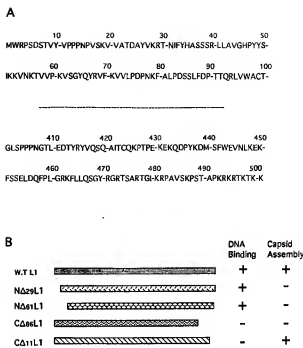

FIG. 7. Summary of L1 proteins analyzed. The L1 amino- and carboxy-terminal protein sequences (A) are shown along with the truncated L1 proteins (B) tested for DNA binding and capsid self-assembly. A consensus bipartite NLS is present at residues 480 to 481 through 492 to 495.

protocol was used found no DNA binding activity for this protein, although the HPV-16 L1 protein has a similar basic region at its carboxy terminus (31). However, fusion proteins for HPV-16 and 6b L1 proteins have demonstrated DNA binding (16); thus, with the current data we believe that L1 is a DNA-binding protein. However, this activity may be modulated if associated with L2 in a manner similar to VP1 association with VP2/VP3 (5). A nuclear localization signal (NLS) sequence was also identified in this region of the HPV-11 L1 protein (29). By inspection, it is likely that a similar NLS sequence may be present in the HPV-11 L1 protein as well (Fig. 7). By analogy to polyomavirus VP1, the DNA binding and NLS sequences may overlap, and further site-directed mutagenesis would be required to identify residues important for each function (18, 19). The biological significance of this DNA binding property also requires further investigation.

The assembly of capsid-like aggregates of HPV-11 L1 par-

allels the behavior of this protein when expressed in insect cells with baculovirus vectors (23). However, the VLPs purified from the nucleus of the insect cells were more homogeneous in appearance than those generated from in vitro assembly of the *E. coli* recombinant L1 protein. This behavior was previously noted for the polyomavirus VP1 protein and has been attributed to more ideal assembly conditions present intracellularly than in vitro (17). In contrast to polyomavirus VP1, the in vitro-assembled L1 is very difficult to disassemble under conditions of calcium chelation and disulfide bond reduction (14a), and by sucrose gradient analysis the majority of the protein was in an aggregated form. Only with concurrent trypsin digestion could capsid-like aggregates be fully disrupted, again suggesting differences in interpentamer bonding between L1 and polyomavirus VP1. The recent observation of "holes" in a subset of BPV capsids suggests an alternative mechanism of capsid dissociation (3). Calcium chelation and/or disulfide bond reduction may cause a conformational change in the capsid, resulting in the opening of holes between capsomeres. The carboxy termini of the L1 proteins may then become exposed to a trypsin-like protease, which cleaves L1, resulting in irreversible capsid disassembly. This type of conformational change with subsequent sensitivity to proteases has been well described for plant viruses (11, 13, 20) and may be utilized during virus uncoating and infection. Without induction of such a conformational change, however, BPV virions are resistant to trypsin and other proteases (21).

The ELISA analysis suggests that HPV-11 L1 purified after expression in *E. coli* is antigenically similar to VLPs and thus to native HPV-11 virions. The results obtained with the trpL1 protein also demonstrate that the unassembled capsomere subunit retains the overall antigenicity of VLPs and capsids, since the reactivities with polyclonal antisera were indistinguishable

between capsomeres (trpL1), capsids, and VLPs over the titration range of the ELISA. Previous attempts to generate native immunoreactivity with L1 proteins purified after expression in *E. coli* may have been unsuccessful because these proteins were denatured. Although the ability of individual capsomeres to induce HPV-11 virus-neutralizing antibodies remains to be tested, these results suggest that L1 capsomeres may possess a neutralizing conformational domain, characterized previously with the same antisera (4, 22, 24).

The ability to purify the HPV-11 L1 protein in a soluble, nondenatured form after expression in *E. coli* now permits a detailed analysis of the intercapsomere bonding domain and buffer conditions which favor papillomavirus capsid assembly. Further studies will be required to test whether the recombinant HPV-11 capsomeres can generate a neutralizing-antibody response which is biologically significant.

REFERENCES

1. Aebersold, R. 1989. Internal amino acid sequence analysis of proteins after in situ protease digestion on nitrocellulose, p. 71–88. *In* P. T. Matsudaira (ed.), A practical guide to protein and peptide purification for microsequencing. Academic Press, San Diego, Calif.
2. Baker, T. S., W. W. Newcomb, N. H. Olson, L. M. Cowsert, C. Olson, and J. C. Brown. 1991. Structures of bovine and human papillomaviruses: analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophys. J. **60:**1445–1456.
3. Belnap, D. M., N. H. Olson, N. M. Cladel, W. W. Newcomb, J. C. Brown, J. W. Kreider, N. D. Christensen, and T. S. Baker. 1996. Conserved features in papillomavirus and polyomavirus capsids. J. Mol. Biol. **259:**249–263.
4. Bonnez, W., R. C. Rose, and R. C. Reichman. 1992. Antibody-mediated neutralization of human papillomavirus type 11 (HPV-11) infection in the nude mouse: detection of HPV-11 mRNAs. J. Infect. Dis. **165:**376–380.
4a. Cripe, T. Unpublished results.
5. Delos, S. E., T. P. Cripe, A. D. Leavitt, H. Greisman, and R. L. Garcea. 1995. Expression of the polyomavirus minor capsid proteins, VP2 and VP3, in *Escherichia coli*: in vitro interactions with recombinant VP1 capsomeres. J. Virol. **69:**7734–7742.
6. Doorbar, J., and P. H. Gallimore. 1987. Identification of proteins encoded by the L1 and L2 open reading frames of human papillomavirus 1a. J. Virol. **61:**2793–2799.
7. Favre, M., F. Breitburd, O. Croissant, and G. Orth. 1975. Structural polypeptides of rabbit, bovine, and human papillomaviruses. J. Virol. **15:**1239–1247.
8. Garcea, R. L., D. M. Salunke, and D. L. D. Caspar. 1987. Site-directed mutation affecting polyomavirus self-assembly in vitro. Nature (London) **329:**86–87.
9. Gharakhanian, E., J. Takahashi, J. Clever, and H. Kasamatsu. 1988. In vitro assay for protein-protein interaction: carboxyl-terminal 40 residues of simian virus 40 structural protein VP3 contain a determinant for interaction with VP1. Proc. Natl. Acad. Sci. USA **85:**6607–6611.
10. Hagensee, M. E., N. Yeagashi, and D. A. Galloway. 1993. Self-assembly of

11. human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. J. Virol. **67:**315–322.
11. Hsu, C. H., O. P. Sehgal, and E. E. Pickett. 1976. Stabilizing effect of divalent metal ions on virions of southern bean mosaic virus. Virology **69:**587–595.
12. Kirnbauer, R., G. Booy, N. Cheng, D. R. Lowy, and J. T. Schiller. 1992. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc. Natl. Acad. Sci. USA **89:**12180–12184.
13. Kruse, J., K. M. Kruse, J. Witz, C. Chauvin, B. Jacrot, and A. Tardieu. 1982. Divalent ion-dependent reversible swelling of tomato bushy stunt virus and organization of the expanded virion. J. Mol. Biol. **162:**393–417.
14. Leavitt, A. D., T. M. Roberts, and R. L. Garcea. 1985. Polyoma virus major capsid protein, VP1: purification after high level expression in *Escherichia coli*. J. Biol. Chem. **260:**12803–12809.
14a. M. Personal communication.
15. Liddington, R. C., Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison. 1991. Structure of simian virus 40 at 3.8-Å resolution. Nature (London) **354:**278–284.
16. Mallon, R. G., D. Wojciechowicz, and V. Defendi. 1987. DNA-binding activity of papillomavirus proteins. J. Virol. **61:**1655–1660.
17. Montross, L., S. Watkins, R. B. Moreland, H. Mamon, D. L. D. Caspar, and R. L. Garcea. 1991. Nuclear assembly of polyomavirus capsids in insect cells expressing the major capsid protein VP1. J. Virol. **65:**4991–4998.
18. Moreland, R. B., and R. L. Garcea. 1991. Characterization of a nuclear localization sequence in the polyomavirus capsid protein VP1. Virology **185:**513–518.
19. Moreland, R. B., L. Montross, and R. L. Garcea. 1991. Characterization of the DNA-binding properties of the polyomavirus capsid protein VP1. J. Virol. **65:**1168–1176.
20. Rayment, I., J. E. Johnson, and M. G. Rossmann. 1979. Metal-free southern bean mosaic virus crystals. J. Biol. Chem. **254:**5243–5245.
21. Roden, R. B. S., N. L. Hubbert, R. Kirnbauer, F. Breitburd, D. R. Lowy, and J. T. Schiller. 1995. Papillomavirus L1 capsids agglutinate mouse erythrocytes through a proteinaceous receptor. J. Virol. **69:**5147–5151.
22. Rose, R. C., W. Bonnez, C. Da Rin, D. J. McCance, and R. C. Reichman. 1994. Serological differentiation of human papillomavirus types 11, 16, and 18 using recombinant virus-like particles. J. Gen. Virol. **75:**2445–2449.
23. Rose, R. C., W. Bonnez, R. C. Reichman, and R. L. Garcea. 1993. Expression of human papillomavirus type 11 L1 protein in insect cells: in vivo and in vitro assembly of viruslike particles. J. Virol. **67:**1936–1944.
24. Rose, R. C., R. C. Reichman, and W. Bonnez. 1994. Human papillomavirus (HPV) type 11 recombinant virus-like particles induce the formation of neutralizing antibodies and detect HPV-specific antibodies in human sera. J. Gen. Virol. **75:**2075–2079.
25. Salunke, D. M., D. L. D. Caspar, and R. L. Garcea. 1986. Self-assembly of purified polyomavirus capsid protein VP1. Cell **46:**895–904.
26. Salunke, D. M., D. L. D. Caspar, and R. L. Garcea. 1989. Polymorphism in the assembly of polyomavirus capsid protein VP1. Biophys. J. **56:**887–900.
26a. Schiller, J. Personal communication.
27. Stehle, T., and S. C. Harrison. 1996. Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. Structure **4:**183–194.
28. Volpers, C., P. Schirmacher, R. E. Streeck, and M. Sapp. 1994. Assembly of the major and the minor capsid protein of human papillomavirus type 33 into virus-like particles and tubular structures in insect cells. Virology **200:**504–512.
29. Zhou, J., J. Doorbar, X.-Y. Sun, L. V. Crawford, C. S. McLean, and I. H. Frazer. 1991. Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. Virology **185:**625–632.
30. Zhou, J., D. J. Stenzel, X. Y. Sun, and I. H. Frazer. 1993. Synthesis and assembly of infectious bovine papillomavirus particles in vitro. J. Gen. Virol. **74:**763–768.
31. Zhou, J., X. Y. Sun, K. Louis, and I. H. Frazer. 1994. Interaction of human papillomavirus (HPV) type 16 capsid proteins with HPV DNA requires an intact L2 N-terminal sequence. J. Virol. **68:**619–625.